GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: March 10, 2003, 17:00:21 ; Search time 7.38462 Seconds
(without alignments)
67.399 Million cell updates/sec

Title: US-09-993-392-1
Perfect score: 66
Sequence: 1 RSPNHIVVLCRG 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	66	100.0	612	1	OCTC_HUMAN	Q9ukg9 homo sapien
2	50	75.8	612	1	OCTC_BOVIN	O19094 bos taurus
3	47	71.2	612	1	OCTC_RAT	P11466 rattus norv
4	42	63.6	314	1	YFCB_HAEIN	P45106 haemophilus
5	41	62.1	151	1	ARG1_CLOPE	Q46172 clostridium
6	40	60.6	262	1	YA86_MYCTU	O53434 mycobacteri
7	40	60.6	345	1	UXA2_BACHD	Q9kez3 bacillus ha
8	39	59.1	519	1	TYR2_HUMAN	P40126 homo sapien
9	38	57.6	103	1	RNS_CEPSI	Q29539 cephalophus
10	37	56.1	92	1	PLM_MOUSE	Q9z239 mus musculu
11	37	56.1	213	1	RB4B_HUMAN	P22750 homo sapien
12	37	56.1	213	1	RB4B_RAT	P51146 rattus norv
13	37	56.1	357	1	GLN1_MAIZE	P38559 zea mays (m
14	37	56.1	425	1	IM44_CAEEL	O02161 caenorhabdi
15	37	56.1	573	1	YKK3_CAEEL	P34280 caenorhabdi
16	37	56.1	780	1	YNL5_CAEEL	Q21955 caenorhabdi
17	37	56.1	1184	1	BIMC_EMENI	P17120 emericella
18	37	56.1	1237	1	DP3A_CHLTR	O84549 chlamydia t
19	37	56.1	1656	1	ATC8_YEAST	Q12674 saccharomyc
20	36	54.5	153	1	RL22_METTH	O26115 methanobact
21	36	54.5	162	1	PHZB_PSEFL	Q51788 pseudomonas
22	36	54.5	281	1	MSP1_PLAFN	P13820 plasmodium
23	36	54.5	352	1	GLN1_DAUCA	O22504 daucus caro
24	36	54.5	355	1	GLN1_PEA	P08282 pisum sativ
25	36	54.5	358	1	GLNA_LACSA	P23712 lactuca sat
26	36	54.5	416	1	CSD_XYLFA	Q9pda6 xylella fas
27	36	54.5	492	1	E1BL_ADE07	P03245 human adeno
28	36	54.5	608	1	RDPO_SCEOB	P19593 scenedesmus
29	36	54.5	698	1	EFG_VIBCH	Q9kuz7 vibrio chol
30	36	54.5	781	1	HELI_HSVSA	Q01014 herpesvirus
31	36	54.5	981	1	EPA3_BRARE	O13146 brachydanio
32	36	54.5	2291	1	RRPB_BEV	P18458 berne virus
33	35	53.0	152	1	RNP_RATRT	Q9wux3 rattus ratt
34	35	53.0	250	1	UPPS_CHLPN	Q9z7y7 chlamydia p
35	35	53.0	310	1	YFCB_ECOLI	P39199 escherichia
36	35	53.0	430	1	SRYD_DROME	P07664 drosophila
37	35	53.0	533	1	GLPT_HUMAN	P57057 homo sapien
38	35	53.0	721	1	CLAT_DROME	P07668 drosophila
39	35	53.0	1069	1	CI10_HUMAN	Q9nzb2 homo sapien
40	35	53.0	1237	1	DP3A_CHLMU	Q9pjj7 chlamydia m
41	35	53.0	1240	1	DP3A_CHLPN	Q9z7n8 chlamydia p
42	35	53.0	4036	1	RRPL_DUGBV	Q66431 dugbe virus
43	34.5	52.3	527	1	MUD2_YEAST	P36084 saccharomyc
44	34	51.5	92	1	PLM_CANFA	P56513 canis famil
45	34	51.5	104	1	RNS_SAITA	Q29538 saiga tatar

ALIGNMENTS

RESULT 1
OCTC_HUMAN
ID OCTC_HUMAN STANDARD; PRT; 612 AA.
AC Q9UKG9; Q9Y6I2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisomal carnitine octanoyltransferase (EC 2.3.1.-) (COT).
GN CROT OR COT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Skin;
RX MEDLINE=99417547; PubMed=10486279;
RA Perdinandusse S., Mulders J., Ijlst L., Denis S., Dacremont G.,
RA Waterham H.R., Wanders R.J.A.;
RT "Molecular cloning and expression of human carnitine
RT octanoyltransferase: evidence for its role in the peroxisomal beta-
RT oxidation of branched-chain fatty acids.";
RL Biochem. Biophys. Res. Commun. 263:213-218(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Kim D.G., Hlubb C.W., Yun J., Mihalik S.J.;
RT "Cloning of the human gene for carnitine octanoyltransferase.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: BETA-OXIDATION OF FATTY ACIDS. THE HIGHEST ACTIVITY
CC CONCERN THE C6 TO C10 CHAIN LENGTH SUBSTRATE. CONVERTS THE END
CC PRODUCT OF PRISTANIC ACID BETA OXIDATION, 4,8-DIMETHYLNONANOYL-
CC COA, TO ITS CORRESPONDING CARNITINE ESTER.
CC -!- PATHWAY: Fatty acid beta-oxidation cycle.
CC -!- SUBUNIT: MONOMER (PROBABLE).
CC -!- SUBCELLULAR LOCATION: Peroxisomal (Potential).
CC -!- SIMILARITY: BELONGS TO THE CARNITINE/CHOLINE ACETYLTRANSFERASE
CC FAMILY.
CC ---

CC ---
DR EMBL; AF168793; AAF03234.1; -.
DR EMBL; AF073770; AAD41654.1; -.
DR Genew; HGNC:2366; CROT.
DR MIM; 606090; -.
DR InterPro; IPR000542; Carn_acyltransf.
DR InterPro; IPR000865; Microbodies_C.
DR Pfam; PF00755; Carn_acyltransf; 1.
DR PROSITE; PS00439; ACYLTRANSF_C_1; 1.
DR PROSITE; PS00440; ACYLTRANSF_C_2; 1.
DR PROSITE; PS00342; MICROBODIES_CTER; FALSE_NEG.

```
KW   Transferase; Acyltransferase; Fatty acid metabolism; Transport;
KW   Peroxisome.
FT   ACT_SITE    327    327       POTENTIAL.
FT   BINDING     505    505       INVOLVED IN SUBSTRATE (CARNITINE)
FT                                BINDING (BY SIMILARITY).
FT   DOMAIN      534    537       POLY-GLY.
FT   SITE        610    612       MICROBODY TARGETING SIGNAL (POTENTIAL).
FT   CONFLICT    144    144       L -> V (IN REF. 2).
FT   CONFLICT    168    168       V -> G (IN REF. 2).
SQ   SEQUENCE   612 AA;  70192 MW;  39C23B769A54A2F3 CRC64;

  Query Match             100.0%;  Score 66;  DB 1;  Length 612;
  Best Local Similarity   100.0%;  Pred. No. 0.0001;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RSPNHIVVLCRG 12
           ||||||||||||
Db     187 RSPNHIVVLCRG 198


RESULT 2
OCTC_BOVIN
ID   OCTC_BOVIN     STANDARD;      PRT;   612 AA.
AC   O19094;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Peroxisomal carnitine octanoyltransferase (EC 2.3.1.-) (COT).
GN   CROT OR COT.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS.
RC   TISSUE=Liver;
RX   MEDLINE=97433288; PubMed=9288928;
RA   Cronin C.N.;
RT   "cDNA cloning, recombinant expression, and site-directed mutagenesis
RT   of bovine liver carnitine octanoyltransferase -- Arg505 binds the
RT   carboxylate group of carnitine.";
RL   Eur. J. Biochem. 247:1029-1037(1997).
CC   -!- FUNCTION: BETA-OXIDATION OF FATTY ACIDS. THE HIGHEST ACTIVITY
CC       CONCERN THE C6 TO C10 CHAIN LENGTH SUBSTRATE.
CC   -!- PATHWAY: Fatty acid beta-oxidation cycle.
CC   -!- SUBUNIT: MONOMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: Peroxisomal (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CARNITINE/CHOLINE ACETYLTRANSFERASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U65745; AAC48758.1; -.
DR   InterPro; IPR000542; Carn_acyltransf.
DR   InterPro; IPR000865; Microbodies_C.
DR   Pfam; PF00755; Carn_acyltransf; 1.
DR   PROSITE; PS00439; ACYLTRANSF_C_1; 1.
DR   PROSITE; PS00440; ACYLTRANSF_C_2; 1.
DR   PROSITE; PS00342; MICROBODIES_CTER; FALSE_NEG.
KW   Transferase; Acyltransferase; Fatty acid metabolism; Transport;
KW   Peroxisome.
FT   ACT_SITE    327    327       POTENTIAL.
FT   BINDING     505    505       INVOLVED IN SUBSTRATE (CARNITINE)
FT                                BINDING.
FT   DOMAIN      534    537       POLY-GLY.
FT   SITE        610    612       MICROBODY TARGETING SIGNAL (POTENTIAL).
FT   MUTAGEN     505    505       R->N:INCREASE OF KM TOWARDS CARNITINE.
SQ   SEQUENCE   612 AA;  70263 MW;  2D5D91A54CF8E2BA CRC64;

  Query Match             75.8%;  Score 50;  DB 1;  Length 612;
  Best Local Similarity   72.7%;  Pred. No. 0.1;
  Matches    8;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       2 SPNHIVVLCRG 12
           ||:|: |||||
Db     188 SPSHLAVLCRG 198


RESULT 3
OCTC_RAT
ID   OCTC_RAT       STANDARD;      PRT;   612 AA.
AC   P11466; P48033;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Peroxisomal carnitine octanoyltransferase (EC 2.3.1.-) (COT).
GN   CROT OR COT.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=96085126; PubMed=7495866;
RA   Choi S.J., Oh D.H., Song C.S., Roy A.K., Chatterjee B.;
RT   "Molecular cloning and sequence analysis of the rat liver carnitine
RT   octanoyltransferase cDNA, its natural gene and the gene promoter.";
RL   Biochim. Biophys. Acta 1264:215-222(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=89166509; PubMed=3233218;
RA   Chatterjee B., Song C.S., Kim J.-M., Roy A.K.;
RT   "Cloning, sequencing, and regulation of rat liver carnitine
RT   octanoyltransferase: transcriptional stimulation of the enzyme during
RT   peroxisome proliferation.";
RL   Biochemistry 27:9000-9006(1988).
CC   -!- FUNCTION: BETA-OXIDATION OF FATTY ACIDS. THE HIGHEST ACTIVITY
CC       CONCERN THE C6 TO C10 CHAIN LENGTH SUBSTRATE.
CC   -!- PATHWAY: Fatty acid beta-oxidation cycle.
CC   -!- SUBCELLULAR LOCATION: Peroxisomal (Potential).
CC   -!- TISSUE SPECIFICITY: LIVER.
CC   -!- SIMILARITY: BELONGS TO THE CARNITINE/CHOLINE ACETYLTRANSFERASE
CC       FAMILY.
CC   -!- CAUTION: REF.1 SEQUENCE WAS INCORRECT, STARTING WITH POSITION 548.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U26033; AAC52317.1; -.
DR   EMBL; J02844; AAA40948.1; ALT_SEQ.
DR   PIR; A31948; A31948.
DR   InterPro; IPR000542; Carn_acyltransf.
DR   Pfam; PF00755; Carn_acyltransf; 1.
DR   PROSITE; PS00342; MICROBODIES_CTER; 1.
DR   PROSITE; PS00439; ACYLTRANSF_C_1; 1.
DR   PROSITE; PS00440; ACYLTRANSF_C_2; 1.
KW   Transferase; Acyltransferase; Fatty acid metabolism; Transport;
KW   Peroxisome.
FT   ACT_SITE    327    327       POTENTIAL.
FT   BINDING     505    505       INVOLVED IN SUBSTRATE (CARNITINE) BINDING
FT                                (BY SIMILARITY).
FT   DOMAIN      534    537       POLY-GLY.
```

```
FT   SITE        610    612       MICROBODY TARGETING SIGNAL (POTENTIAL).
FT   CONFLICT    334    334       L -> F (IN REF. 2).
FT   CONFLICT    463    464       VR -> RQ (IN REF. 2).
SQ   SEQUENCE   612 AA;  70302 MW;  41B2F3474C8838D1 CRC64;

  Query Match             71.2%;  Score 47;  DB 1;  Length 612;
  Best Local Similarity   80.0%;  Pred. No. 0.36;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 PNHIVVLCRG 12
          | || |||||
Db    189 PTHIAVLCRG 198


RESULT 4
YFCB_HAEIN
ID   YFCB_HAEIN     STANDARD;      PRT;   314 AA.
AC   P45106;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical adenine-specific methylase HI1201 (EC 2.1.1.72).
GN   HI1201.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC       adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC   -!- SIMILARITY: BELONGS TO THE N6-METHYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32799; AAC22855.1; -.
DR   TIGR; HI1201; -.
DR   InterPro; IPR004556; HemK.
DR   InterPro; IPR002052; N6_Mtase.
DR   InterPro; IPR000051; SAM_bind.
DR   TIGRFAMs; TIGR00536; hemK_fam; 1.
DR   PROSITE; PS00092; N6_MTASE; 1.
KW   Hypothetical protein; Transferase; Methyltransferase;
KW   Complete proteome.
SQ   SEQUENCE   314 AA;  35590 MW;  2AC862003FE05301 CRC64;

  Query Match             63.6%;  Score 42;  DB 1;  Length 314;
  Best Local Similarity   58.3%;  Pred. No. 1.6;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 RSPNHIVVLCRG 12
          : |||| : || |
Db    140 QEPNHILDLCTG 151


RESULT 5
ARG1_CLOPE
ID   ARG1_CLOPE     STANDARD;      PRT;   151 AA.
AC   Q46172;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Arginine repressor 1.
GN   ARGR1 OR ARGR OR AHRC OR CPE0172.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   MEDLINE=97199138; PubMed=9053381;
RA   Ohtani K., Bando M., Swe T., Banu S., Oe M., Hayashi H., Shimizu T.;
RT   "Collagenase gene (colA) is located in the 3'-flanking region of the
RT   perfringolysin O (pfoA) locus in Clostridium perfringens.";
RL   FEMS Microbiol. Lett. 146:155-159(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC   -!- FUNCTION: PUTATIVE REGULATOR FOR ARCABDC OPERON.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE ARGR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X97768; CAA66368.1; -.
DR   EMBL; AP003185; BAB79878.1; -.
DR   HSSP; O31408; 1B4A.
DR   InterPro; IPR001669; Arg_repress.
DR   Pfam; PF01316; Arg_repressor; 1.
DR   ProDom; PD007402; Arg_repress; 1.
KW   Transcription regulation; DNA-binding; Trans-acting factor; Repressor;
KW   Complete proteome.
FT   CONFLICT    126    126       K -> T (IN REF. 1).
FT   CONFLICT    144    151       KELDSLRV -> RN (IN REF. 1).
SQ   SEQUENCE   151 AA;  17427 MW;  9A9D411E0E4C9A9C CRC64;

  Query Match             62.1%;  Score 41;  DB 1;  Length 151;
  Best Local Similarity   87.5%;  Pred. No. 1.1;
  Matches    7;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 NHIVVLCR 11
          |||:||||
Db    127 NHILVLCR 134


RESULT 6
YA86_MYCTU
ID   YA86_MYCTU     STANDARD;      PRT;   262 AA.
AC   O53434;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein Rv1086.
```

```
GN   RV1086 OR MT1118 OR MTV017.39.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE UPP SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021897; CAA17202.1; -.
DR   EMBL; AE006992; AAK45374.1; -.
DR   TIGR; MT1118; -.
DR   TubercuList; Rv1086; -.
DR   InterPro; IPR001441; UPP_synth.
DR   Pfam; PF01255; UPP_synthetase; 1.
DR   ProDom; PD003461; UPP_synth; 1.
DR   TIGRFAMs; TIGR00055; uppS; 1.
DR   PROSITE; PS01066; UPP_SYNTHETASE; 1.
KW   Hypothetical protein; Transferase; Complete proteome.
SQ   SEQUENCE   262 AA;  29410 MW;  2D6745748FE22518 CRC64;

  Query Match             60.6%;  Score 40;  DB 1;  Length 262;
  Best Local Similarity   70.0%;  Pred. No. 3.1;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 PNHIVVLCRG 12
           | || ||| |
Db      32 PRHIAVLCDG 41


RESULT 7
UXA2_BACHD
ID   UXA2_BACHD     STANDARD;      PRT;   345 AA.
AC   Q9KEZ3;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Mannonate dehydratase 2 (EC 4.2.1.8) (D-mannonate hydrolase 2).
GN   UXUA2 OR BH0706.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
CC   -!- CATALYTIC ACTIVITY: D-mannonate = 2-dehydro-3-deoxy-D-gluconate +
CC       H(2)O.
CC   -!- PATHWAY: Glucuronate pathway.
CC   -!- SIMILARITY: BELONGS TO THE MANNONATE DEHYDRATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP001509; BAB04425.1; -.
KW   Lyase; Complete proteome.
SQ   SEQUENCE   345 AA;  38888 MW;  E219AA943820EBEA CRC64;

  Query Match             60.6%;  Score 40;  DB 1;  Length 345;
  Best Local Similarity   54.5%;  Pred. No. 4.1;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 SPNHIVVLCRG 12
             |||| : :| |
Db     228 SPNHGITMCSG 238


RESULT 8
TYR2_HUMAN
ID   TYR2_HUMAN     STANDARD;      PRT;   519 AA.
AC   P40126;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dopachrome tautomerase precursor (EC 5.3.3.12) (DT) (DCT) (Dopachrome
DE   delta-isomerase) (Tyrosinase-related protein 2) (TRP-2) (TRP2).
GN   DCT OR TYRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94198295; PubMed=8148378;
RA   Yokoyama K., Suzuki H., Yasumoto K.I., Tomita Y., Shibahara S.;
RT   "Molecular cloning and functional analysis of a cDNA coding for human
RT   DOPAchrome tautomerase/tyrosinase-related protein-2.";
RL   Biochim. Biophys. Acta 1217:317-321(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94266170; PubMed=8206391;
RA   Cassady J.L., Sturm R.A.;
RT   "Sequence of the human dopachrome tautomerase-encoding TRP-2 cDNA.";
RL   Gene 143:295-298(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94139684; PubMed=8306979;
RA   Bouchard B., del Marmol V., Jackson I.J., Cherif D., Dubertret L.;
RT   "Molecular characterization of a human tyrosinase-related-protein-2
RT   cDNA. Patterns of expression in melanocytic cells.";
RL   Eur. J. Biochem. 219:127-134(1994).
RN   [4]
RP   SEQUENCE OF 1-98 FROM N.A.
```

```
RC   TISSUE=Liver;
RX   MEDLINE=96079088; PubMed=8530077;
RA   Sturm R.A., O'Sullivan B.J., Box N.F., Smith A.G., Smit S.E.,
RA   Puttick E.R.J., Parsons P.G., Dunn I.S.;
RT   "Chromosomal structure of the human TYRP1 and TYRP2 loci and
RT   comparison of the tyrosinase-related protein gene family.";
RL   Genomics 29:24-34(1995).
RN   [5]
RP   SEQUENCE OF 1-98 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95014579; PubMed=7929451;
RA   Yokoyama K., Yasumoto K.I., Suzuki H., Shibahara S.;
RT   "Cloning of the human DOPAchrome tautomerase/tyrosinase-related
RT   protein 2 gene and identification of two regulatory regions required
RT   for its pigment cell-specific expression.";
RL   J. Biol. Chem. 269:27080-27087(1994).
CC   -!- FUNCTION: INVOLVED IN REGULATING EUMELANIN AND PHAEOMELANIN
CC       LEVELS.
CC   -!- CATALYTIC ACTIVITY: L-dopachrome = 5,6-dihydroxyindole-2-
CC       carboxylate.
CC   -!- COFACTOR: Binds 2 zinc ions (By similarity).
CC   -!- PATHWAY: Melanin biosynthesis.
CC   -!- SUBUNIT: TYROSINASE, TYRP1 AND TYRP2 MAY FORM A MULTIENZYME
CC       COMPLEX (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Melanosomal.
CC   -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17547; BAA04484.1; -.
DR   EMBL; L18967; AAA20870.1; -.
DR   EMBL; S69231; AAC60627.1; -.
DR   EMBL; L38953; AAC41925.1; -.
DR   EMBL; D28767; BAA05956.1; -.
DR   PIR; S43510; S43510.
DR   Genew; HGNC:2709; DCT.
DR   MIM; 191275; -.
DR   InterPro; IPR002227; Tyrosinase.
DR   Pfam; PF00264; tyrosinase; 1.
DR   PRINTS; PR00092; TYROSINASE.
DR   PROSITE; PS00497; TYROSINASE_1; 1.
DR   PROSITE; PS00498; TYROSINASE_2; 1.
KW   Isomerase; Zinc; Glycoprotein; Signal; Transmembrane;
KW   Melanin biosynthesis.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24    519       DOPACHROME TAUTOMERASE.
FT   DOMAIN      24    472       LUMENAL, MELANOSOME (POTENTIAL).
FT   TRANSMEM   473    493       POTENTIAL.
FT   DOMAIN     494    519       CYTOPLASMIC (POTENTIAL).
FT   METAL      189    189       ZINC A (BY SIMILARITY).
FT   METAL      211    211       ZINC A (BY SIMILARITY).
FT   METAL      220    220       ZINC A (BY SIMILARITY).
FT   METAL      369    369       ZINC B (BY SIMILARITY).
FT   METAL      373    373       ZINC B (BY SIMILARITY).
FT   METAL      396    396       ZINC B (BY SIMILARITY).
FT   CARBOHYD   170    170       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   237    237       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   300    300       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   342    342       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   377    377       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   519 AA;  59145 MW;  AFDDF21768002A89 CRC64;

  Query Match             59.1%;  Score 39;  DB 1;  Length 519;
  Best Local Similarity   66.7%;  Pred. No. 9.6;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 NHIVVLCRG 12
           ||:| || |
Db     293 NHLVTLCNG 301


RESULT 9
RNS_CEPSI
ID   RNS_CEPSI      STANDARD;      PRT;   103 AA.
AC   Q29539;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ribonuclease, seminal (EC 3.1.27.5) (Seminal RNase) (Fragment).
GN   SRN.
OS   Cephalophus silvicultor (Yellow-backed duiker).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Cephalophinae; Cephalophus.
OX   NCBI_TaxID=50347;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96184512; PubMed=8605993;
RA   Trabesinger-Ruef N., Jermann T., Zankel T., Durrant B., Frank G.,
RA   Benner S.A.;
RT   "Pseudogenes in ribonuclease evolution: a source of new
RT   biomacromolecular function?";
RL   FEBS Lett. 382:319-322(1996).
CC   -!- FUNCTION: THIS ENZYME HYDROLYZES BOTH SINGLE- AND DOUBLE-STRANDED
CC       RNA.
CC   -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to nucleoside 3'-
CC       phosphates and 3'-phosphooligonucleotides ending in C-P or U-P
CC       with 2',3'-cyclic phosphate intermediates.
CC   -!- SUBUNIT: HOMODIMER; DISULFIDE-LINKED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PANCREATIC RIBONUCLEASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S81529; AAB39847.1; -.
DR   HSSP; P00669; 1BSR.
DR   InterPro; IPR001427; RNaseA.
DR   Pfam; PF00074; rnaseA; 1.
DR   ProDom; PD000535; RNaseA; 1.
DR   SMART; SM00092; RNAse_Pc; 1.
DR   PROSITE; PS00127; RNASE_PANCREATIC; 1.
KW   Hydrolase; Nuclease; Endonuclease.
FT   NON_TER      1      1
FT   ACT_SITE    27     27       BY SIMILARITY.
FT   DISULFID    12     70       BY SIMILARITY.
FT   DISULFID    26     81       BY SIMILARITY.
FT   DISULFID    44     96       BY SIMILARITY.
FT   DISULFID    51     58       BY SIMILARITY.
FT   DISULFID    18     18       INTERCHAIN (BY SIMILARITY).
FT   NON_TER    103    103
SQ   SEQUENCE   103 AA;  11357 MW;  6F8AE8FDE8957DF1 CRC64;

  Query Match             57.6%;  Score 38;  DB 1;  Length 103;
  Best Local Similarity   50.0%;  Pred. No. 2.7;
  Matches    6;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 RSPNHIVVLCRG 12
           |:  ||:| | |
Db      87 RAEKHIIVACEG 98


RESULT 10
PLM_MOUSE
ID   PLM_MOUSE      STANDARD;      PRT;    92 AA.
```

```
AC   Q9Z239;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Phospholemman precursor (FXYD domain-containing ion transport
DE   regulator 1).
GN   FXYD1 OR PLM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RA   Bogaev R.C., Kobayashi Y.M., Mounsey J.P., Moorman J.R., Jones L.R.,
RA   Tucker A.L.;
RT   "Gene structure and expression of phospholemman in mouse.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- FUNCTION: INDUCES A HYPERPOLARIZATION-ACTIVATED CHLORIDE CURRENT
CC       WHEN EXPRESSED IN XENOPUS OOCYTES. MAY HAVE A FUNCTIONAL ROLE IN
CC       MUSCLE CONTRACTION (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: MAJOR PLASMA MEMBRANE SUBSTRATE FOR CAMP-DEPENDENT PROTEIN
CC       KINASE (PK-A) AND PROTEIN KINASE C (PK-C) IN SEVERAL DIFFERENT
CC       TISSUES. PHOSPHORYLATED IN RESPONSE TO INSULIN AND ADRENERGIC
CC       STIMULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FXYD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF091390; AAD11781.1; -.
DR   EMBL; AF089734; AAD41683.1; -.
DR   EMBL; AK002585; BAB22208.1; -.
DR   MGD; MGI:1889273; Fxyd1.
DR   InterPro; IPR000272; ATP1G1_PLM_MAT8.
DR   Pfam; PF02038; ATP1G1_PLM_MAT8; 1.
DR   PROSITE; PS01310; FXYD; 1.
KW   Transmembrane; Phosphorylation; Signal; Ionic channel; Ion transport.
FT   SIGNAL       1     20       BY SIMILARITY.
FT   CHAIN       21     92       PHOSPHOLEMMAN.
FT   DOMAIN      21     35       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    36     56       POTENTIAL.
FT   DOMAIN      57     92       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     83     83       PHOSPHORYLATION (BY PKC AND PKA) (BY
FT                               SIMILARITY).
FT   MOD_RES     88     88       PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
SQ   SEQUENCE   92 AA;  10323 MW;  0BDB1DC83417F3AD CRC64;

  Query Match             56.1%;  Score 37;  DB 1;  Length 92;
  Best Local Similarity   66.7%;  Pred. No. 3.8;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 SPNHIVVLC 10
           || ||: ||
Db       3 SPGHILALC 11


RESULT 11
RB4B_HUMAN
ID   RB4B_HUMAN     STANDARD;      PRT;   213 AA.
AC   P22750;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ras-related protein Rab-4B.
GN   RAB4B.
OS   Homo sapiens (Human), and
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606, 9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Human;
RA   Huang C., Wu T., Xu S., Gu W., Wang Y., Han Z., Chen Z.;
RT   "Novel genes expressed in hematopoietic stem/progenitor cells from
RT   myelodysplastic syndromes patient.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=C.familiaris; STRAIN=Cocker spaniel;
RX   MEDLINE=91061765; PubMed=2123294;
RA   Chavrier P., Vingron M., Sander C., Simons K., Zerial M.;
RT   "Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell
RT   line.";
RL   Mol. Cell. Biol. 10:6578-6585(1990).
CC   -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC       TRAFFIC (BY SIMILARITY).
CC   -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF165522; AAD45923.1; -.
DR   EMBL; X56389; CAA39800.1; -.
DR   PIR; F36364; F36364.
DR   HSSP; P36017; 1EK0.
DR   Genew; HGNC:9782; RAB4B.
DR   InterPro; IPR003579; GTPase_Rab.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR005225; Small_GTP.
DR   Pfam; PF00071; ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   SMART; SM00175; RAB; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
KW   GTP-binding; Lipoprotein; Prenylation; Protein transport.
FT   NP_BIND     15     22       GTP (BY SIMILARITY).
FT   NP_BIND     63     67       GTP (BY SIMILARITY).
FT   NP_BIND    121    124       GTP (BY SIMILARITY).
FT   DOMAIN      37     45       EFFECTOR REGION (BY SIMILARITY).
FT   LIPID      211    211       GERANYL-GERANYL (BY SIMILARITY).
```

```
FT   LIPID       213    213       GERANYL-GERANYL (BY SIMILARITY).
SQ   SEQUENCE   213 AA;  23587 MW;  0C3D76DC3285DB98 CRC64;

  Query Match             56.1%;  Score 37;  DB 1;  Length 213;
  Best Local Similarity   66.7%;  Pred. No. 9;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 SPNHIVVLC 10
          ||| :|:||
Db    111 SPNIVVILC 119


RESULT 12
RB4B_RAT
ID   RB4B_RAT       STANDARD;      PRT;   213 AA.
AC   P51146;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Ras-related protein Rab-4B.
GN   RAB4B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Heart muscle;
RA   Schuermann A., Muehl-Zuerbes P., Lie C., Joost H.G.;
RL   Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
CC       TRAFFIC (BY SIMILARITY).
CC   -!- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78605; CAA55339.1; -.
DR   HSSP; P36017; 1EK0.
DR   InterPro; IPR003579; GTPase_Rab.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR005225; Small_GTP.
DR   Pfam; PF00071; ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   SMART; SM00175; RAB; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
KW   GTP-binding; Lipoprotein; Prenylation; Protein transport.
FT   NP_BIND      15     22       GTP (BY SIMILARITY).
FT   NP_BIND      63     67       GTP (BY SIMILARITY).
FT   NP_BIND     121    124       GTP (BY SIMILARITY).
FT   DOMAIN       37     45       EFFECTOR REGION (BY SIMILARITY).
FT   LIPID       211    211       GERANYL-GERANYL (BY SIMILARITY).
FT   LIPID       213    213       GERANYL-GERANYL (BY SIMILARITY).
SQ   SEQUENCE   213 AA;  23629 MW;  0C3D76DC328B0018 CRC64;

  Query Match             56.1%;  Score 37;  DB 1;  Length 213;
  Best Local Similarity   66.7%;  Pred. No. 9;
  Matches    6;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 SPNHIVVLC 10
          ||| :|:||
Db    111 SPNIVVILC 119


RESULT 13
GLN1_MAIZE
ID   GLN1_MAIZE     STANDARD;      PRT;   357 AA.
AC   P38559;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Glutamine synthetase root isozyme 1 (EC 6.3.1.2) (Glutamate--ammonia
DE   ligase) (GS122).
GN   GLN6 OR GS1-1.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. A188; TISSUE=Seedling;
RX   MEDLINE=94033318; PubMed=8106013;
RA   Li M.-G., Villemur R., Hussey P.J., Silflow C.D., Gantt J.S.,
RA   Snustad D.P.;
RT   "Differential expression of six glutamine synthetase genes in Zea
RT   mays.";
RL   Plant Mol. Biol. 23:401-407(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Golden cross Bantam T51; TISSUE=Leaf;
RA   Sakakibara H., Kawabata S., Takahashi H., Hase T., Sugiyama T.;
RT   "Molecular cloning of the family of glutamine synthetase genes from
RT   maize: expression of genes for glutamine synthetase and ferredoxin-
RT   dependent glutamate synthase in photosynthetic and non-photosynthetic
RT   tissues.";
RL   Plant Cell Physiol. 33:49-58(1992).
RN   [3]
RP   REVISIONS.
RA   Sakakibara H.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PLAYS A ROLE IN THE FLOW OF NITROGEN INTO NITROGENOUS
CC       ORGANIC COMPOUNDS.
CC   -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC       L-glutamine.
CC   -!- SUBUNIT: HOMOOCTAMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: FOUND MAINLY IN THE CORTICAL TISSUES OF
CC       SEEDLING ROOTS, AND IN THE ROOT TIP.
CC   -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X65926; CAA46719.1; -.
DR   EMBL; D14579; BAA03433.1; -.
DR   PIR; S39477; S39477.
DR   MaizeDB; 17151; -.
DR   InterPro; IPR001691; GLN_synth.
DR   InterPro; IPR001637; GlnA_adenyltn.
DR   Pfam; PF00120; gln-synt; 1.
DR   PROSITE; PS00180; GLNA_1; 1.
DR   PROSITE; PS00181; GLNA_ATP; 1.
KW   Ligase; Multigene family.
FT   CONFLICT     48     48       I -> S (IN REF. 2).
SQ   SEQUENCE   357 AA;  39250 MW;  912A5E3BAF9CC2B8 CRC64;

  Query Match             56.1%;  Score 37;  DB 1;  Length 357;
  Best Local Similarity   60.0%;  Pred. No. 15;
  Matches    6;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 RSPNHIVVLC 10
          |  |||:|:|
Db     83 RKGNHILVMC 92
```

```
RESULT 14
IM44_CAEEL
ID   IM44_CAEEL     STANDARD;      PRT;   425 AA.
AC   O02161;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable import inner membrane translocase subunit TIM44,
DE   mitochondrial precursor.
GN   T09B4.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Langston Z., Wohldmann P., Gillam B.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
CC       PROBABLY INVOLVED IN TRANSLOCATION ACROSS THE INNER MEMBRANE. AS A
CC       BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF PREPROTEINS.
CC       RECRUITS MITOCHONDRIAL HSP70 TO DRIVE PROTEIN TRANSLOCATION INTO
CC       THE MATRIX USING ATP AS AN ENERGY SOURCE (BY SIMILARITY).
CC   -!- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC       LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44) (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane (Potential).
CC   -!- SIMILARITY: BELONGS TO THE TIM44 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U97405; AAB53011.1; -.
DR   WormPep; T09B4.9; CE13473.
DR   TIGRFAMs; TIGR00984; 3a0801s03tim44; 1.
KW   Mitochondrion; Inner membrane; Transport; Protein transport;
KW   Translocation; Transit peptide.
FT   TRANSIT       1      ?       MITOCHONDRION.
FT   CHAIN         ?    425       PROBABLE IMPORT INNER MEMBRANE
FT                                TRANSLOCASE SUBUNIT TIM44.
SQ   SEQUENCE   425 AA;  49398 MW;  203DFBD614E099F8 CRC64;

  Query Match             56.1%;  Score 37;  DB 1;  Length 425;
  Best Local Similarity   87.5%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       4 NHIVVLCR 11
           ||| |||| |
Db     393 NHIWVLCR 400


RESULT 15
YKK3_CAEEL
ID   YKK3_CAEEL     STANDARD;      PRT;   573 AA.
AC   P34280;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical GTP-binding protein C02F5.3 in chromosome III.
GN   C02F5.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -!- SIMILARITY: BELONGS TO THE GTP1 / OBG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14745; AAA27918.1; -.
DR   PIR; S44605; S44605.
DR   WormPep; C02F5.3; CE00039.
DR   InterPro; IPR000765; GTP1_OBG.
DR   InterPro; IPR005225; Small_GTP.
DR   InterPro; IPR004095; TGS_dom.
DR   Pfam; PF01018; GTP1_OBG; 1.
DR   Pfam; PF02824; TGS; 1.
DR   PRINTS; PR00326; GTP1OBG.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS00905; GTP1_OBG; 1.
KW   Hypothetical protein; GTP-binding.
FT   NP_BIND      69     76       GTP (BY SIMILARITY).
FT   NP_BIND     115    119       GTP (BY SIMILARITY).
FT   NP_BIND     246    249       GTP (BY SIMILARITY).
SQ   SEQUENCE   573 AA;  64299 MW;  BA437D93C898B9AC CRC64;

  Query Match             56.1%;  Score 37;  DB 1;  Length 573;
  Best Local Similarity   50.0%;  Pred. No. 25;
  Matches    5;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RSPNHIVVLC 10
           | |:|:|: |
Db     261 RMPHHVVISC 270


Search completed: March 10, 2003, 17:13:49
Job time : 9.38462 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: March 10, 2003, 17:00:21 ; Search time 8.61539 Seconds
(without alignments)
67.399 Million cell updates/sec

Title: US-09-993-392-3
Perfect score: 79
Sequence: 1 RTAAHPAQRRPWRA 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	54.4	107	1	Y149_NPVAC	P41706 autographa
2	43	54.4	239	1	VV_NDVU2	Q06428 newcastle d
3	42	53.2	393	1	DCUP_MAIZE	O81220 zea mays (m
4	41	51.9	286	1	FPG_STRCO	Q9zbq6 streptomyce
5	41	51.9	436	1	EPE2_RALSO	Q45411 ralstonia s
6	41	51.9	2161	1	SHK1_HUMAN	Q9y566 homo sapien
7	40	50.6	242	1	THYL_HUMAN	P20396 homo sapien
8	40	50.6	465	1	Y093_RHIME	O87394 rhizobium m
9	40	50.6	599	1	HMW3_MYCGE	Q57081 mycoplasma
10	39	49.4	55	1	YPU3_RHOCA	P26159 rhodobacter
11	39	49.4	294	1	DRAG_RHORU	P14300 rhodospiril
12	39	49.4	313	1	PENA_BURCE	Q02940 burkholderi
13	39	49.4	314	1	APBE_CHLPN	Q9z8k2 chlamydia p
14	39	49.4	1964	1	NTC4_MOUSE	P31695 mus musculu
15	39	49.4	2003	1	NTC4_HUMAN	Q99466 homo sapien
16	38.5	48.7	277	1	CLPP_HUMAN	Q16740 homo sapien
17	38	48.1	143	1	YDRA_RHORU	P14301 rhodospiril
18	38	48.1	177	1	IL1X_PIG	Q29056 sus scrofa
19	38	48.1	260	1	HXC9_HUMAN	P31274 homo sapien
20	38	48.1	260	1	HXC9_MOUSE	P09633 mus musculu
21	38	48.1	291	1	Y126_TREPA	O83163 treponema p
22	38	48.1	496	1	ADA_MYCTU	Q10630 mycobacteri
23	38	48.1	693	1	RECG_ECO57	Q8xd86 escherichia
24	38	48.1	693	1	RECG_ECOLI	P24230 escherichia
25	38	48.1	1061	1	ANPA_HUMAN	P16066 homo sapien
26	38	48.1	1353	1	CYA9_HUMAN	O60503 homo sapien
27	38	48.1	2109	1	PKS1_ASPPA	Q12053 aspergillus
28	37	46.8	142	1	RADC_COXBU	O85403 coxiella bu
29	37	46.8	351	1	DJB2_HUMAN	P25686 homo sapien
30	37	46.8	542	1	KCCB_MOUSE	P28652 mus musculu
31	37	46.8	542	1	KCCB_RAT	P08413 rattus norv
32	37	46.8	664	1	KCCB_HUMAN	Q13554 homo sapien
33	37	46.8	784	1	KL68_DROME	P46867 drosophila
34	37	46.8	799	1	AFSK_STRCO	P54741 streptomyce
35	37	46.8	992	1	EBN6_EBV	P03204 epstein-bar
36	36.5	46.2	241	1	CRTA_RHOCA	P17055 rhodobacter
37	36.5	46.2	1516	1	CA1H_HUMAN	P39060 homo sapien
38	36	45.6	39	1	PRT2_BUFJA	P24642 bufo japoni
39	36	45.6	138	1	YJ99_ARCFU	O28280 archaeoglob
40	36	45.6	171	1	PAT_STRCO	P21861 streptomyce
41	36	45.6	178	1	IL1X_RAT	P25086 rattus norv
42	36	45.6	199	1	NUOC_RHOCA	O84971 rhodobacter
43	36	45.6	236	1	GTX2_MAIZE	P50472 zea mays (m
44	36	45.6	361	1	IE63_PRVKA	Q85232 pseudorabie
45	36	45.6	508	1	CDPK_SOYBN	P28583 glycine max

ALIGNMENTS

RESULT 1
Y149_NPVAC
ID Y149_NPVAC STANDARD; PRT; 107 AA.
AC P41706;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 12.4 kDa protein in IE1-IEN intergenic region.
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.";
RL Virology 202:586-605(1994).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC ---
DR EMBL; L22858; AAA66779.1; -.
KW Hypothetical protein.
SQ SEQUENCE 107 AA; 12419 MW; 644847D2800FC9AA CRC64;

Query Match 54.4%; Score 43; DB 1; Length 107;
Best Local Similarity 70.0%; Pred. No. 2.9;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      5 HPAQRRPWRA 14
          || || || |
Db     29 HPYQRAPWSA 38
```

RESULT 2
VV_NDVU2
ID VV_NDVU2 STANDARD; PRT; 239 AA.
AC Q06428;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonstructural protein V.
GN P/V.
OS Newcastle disease virus (strain Ulster/2C) (NDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX NCBI_TaxID=36411;
RN [1]

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94103756; PubMed=8277263;
RA   Steward M., Vipond I.B., Millar N.S., Emmerson P.T.;
RT   "RNA editing in Newcastle disease virus.";
RL   J. Gen. Virol. 74:2539-2547(1993).
CC   -!- ALTERNATIVE PRODUCTS: THE P PROTEIN IS COTERMINAL WITH THE V
CC       PROTEIN, AND IS ENCODED BY A SEPARATE MRNA WHICH DIFFERS BY TWO
CC       NONTEMPLATED NUCLEOTIDES. THE SEQUENCE OF P AND V ARE IDENTICAL
CC       UP TO RESIDUE 164.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z26249; -; NOT_ANNOTATED_CDS.
DR   InterPro; IPR004897; Paramyx_P_V.
DR   Pfam; PF03210; Paramyx_P_V; 1.
KW   Nonstructural protein; RNA editing.
FT   DOMAIN      196    224       CYS-RICH.
SQ   SEQUENCE   239 AA;  25379 MW;  E60075ADDAC40DE6 CRC64;

  Query Match             54.4%;  Score 43;  DB 1;  Length 239;
  Best Local Similarity   53.8%;  Pred. No. 6.2;
  Matches    7;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWR 13
           |  | | | |||:
Db     162 RETAEPGQGRPWK 174


RESULT 3
DCUP_MAIZE
ID   DCUP_MAIZE     STANDARD;      PRT;   393 AA.
AC   O81220;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Uroporphyrinogen decarboxylase, chloroplast precursor (EC 4.1.1.37)
DE   (UPD).
GN   LES22.
OS   Zea mays (Maize).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98334526; PubMed=9668130;
RA   Hu G., Yalpani N., Briggs S.P., Johal G.S.;
RT   "A porphyrin pathway impairment is responsible for the phenotype of a
RT   dominant disease lesion mimic mutant of maize.";
RL   Plant Cell 10:1095-1105(1998).
CC   -!- CATALYTIC ACTIVITY: Uroporphyrinogen-III = coproporphyrinogen + 4
CC       CO(2).
CC   -!- PATHWAY: Porphyrin biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Chloroplast.
CC   -!- SIMILARITY: BELONGS TO THE UROPORPHYRINOGEN DECARBOXYLASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF058763; AAC31883.1; -.
DR   HSSP; P06132; 1URO.
DR   InterPro; IPR000257; Uro_decarbxyls.
DR   Pfam; PF01208; URO-D; 1.
DR   PROSITE; PS00906; UROD_1; 1.
DR   PROSITE; PS00907; UROD_2; 1.
KW   Lyase; Decarboxylase; Porphyrin biosynthesis;
KW   Chlorophyll biosynthesis; Chloroplast; Transit peptide.
FT   TRANSIT       1      ?       CHLOROPLAST (POTENTIAL).
FT   CHAIN         ?    393       UROPORPHYRINOGEN DECARBOXYLASE.
SQ   SEQUENCE   393 AA;  43396 MW;  4B419108191668CAB CRC64;

  Query Match             53.2%;  Score 42;  DB 1;  Length 393;
  Best Local Similarity   58.3%;  Pred. No. 14;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TAAHPAQRRPWR 13
           :|| |::|| ||
Db      33 SAAAPSERRSWR 44


RESULT 4
FPG_STRCO
ID   FPG_STRCO      STANDARD;      PRT;   286 AA.
AC   Q9ZBQ6;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Formamidopyrimidine-DNA glycosylase (EC 3.2.2.23) (FAPY-DNA
DE   glycosylase).
GN   MUTM OR FPG OR SCO5573 OR SC7A1.17.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- FUNCTION: THIS ENZYME MAY PLAY A SIGNIFICANT ROLE IN PROCESSES
CC       LEADING TO RECOVERY FROM MUTAGENESIS AND/OR CELL DEATH BY
CC       ALKYLATING AGENTS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of DNA containing ring-opened N(7)-
CC       methylguanine residues, releasing 2,6-diamino-4-hydroxy-5-(N-
CC       methyl)formamidopyrimide.
CC   -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FPG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL034447; CAA22416.1; -.
DR   HSSP; O50606; 1EE8.
DR   InterPro; IPR000191; Fapy_DNA_glyco.
DR   InterPro; IPR000214; Fapy_DNAglyco_zn.
DR   Pfam; PF01149; Fapy_DNA_glyco; 1.
DR   ProDom; PD003680; Fapy_DNA_glyco; 1.
DR   TIGRFAMs; TIGR00577; fpg; 1.
DR   PROSITE; PS01242; FPG; 1.
```

```
KW   DNA repair; Hydrolase; Glycosidase; Zinc; Zinc-finger;
KW   Complete proteome.
FT   ZN_FING     255    278       POTENTIAL.
SQ   SEQUENCE   286 AA;  32511 MW;  C98F08045A9F386B CRC64;

  Query Match             51.9%;  Score 41;  DB 1;  Length 286;
  Best Local Similarity   58.3%;  Pred. No. 15;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 RTAAHPAQRRPW 12
          |  | | :||||
Db    256 RRCATPMRRRPW 267


RESULT 5
EPE2_RALSO
ID   EPE2_RALSO     STANDARD;      PRT;   436 AA.
AC   Q45411;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   EPS I polysaccharide export inner membrane protein epsE.
GN   EPSE.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AW;
RX   MEDLINE=96059643; PubMed=7476194;
RA   Huang J., Schell M.;
RT   "Molecular characterization of the eps gene cluster of Pseudomonas
RT   solanacearum and its transcriptional regulation at a single
RT   promoter.";
RL   Mol. Microbiol. 16:977-989(1995).
CC   -!- FUNCTION: PROBABLY INVOLVED IN POLYMERIZATION AND/OR EXPORT OF
CC       EXOPOLYSACCHARIDE EPS I WHICH FUNCTIONS AS A VIRULENCE FACTOR. MAY
CC       PLAY A ROLE IN EXPORT OF EPS I OR ITS INTERMEDIATES ACROSS THE
CC       MEMBRANES.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: SOME, TO E.COLI BICYCLOMYCIN RESISTANCE PROTEIN (BCR).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U17898; AAA91628.1; -.
KW   Polysaccharide transport; Transport; Transmembrane; Inner membrane.
FT   TRANSMEM     20     40       POTENTIAL.
FT   TRANSMEM     49     69       POTENTIAL.
FT   TRANSMEM     91    111       POTENTIAL.
FT   TRANSMEM    133    153       POTENTIAL.
FT   TRANSMEM    160    180       POTENTIAL.
FT   TRANSMEM    185    205       POTENTIAL.
FT   TRANSMEM    234    254       POTENTIAL.
FT   TRANSMEM    261    281       POTENTIAL.
FT   TRANSMEM    307    327       POTENTIAL.
FT   TRANSMEM    341    361       POTENTIAL.
FT   TRANSMEM    375    395       POTENTIAL.
FT   TRANSMEM    396    416       POTENTIAL.
SQ   SEQUENCE   436 AA;  45765 MW;  27B592A0155A0B04 CRC64;

  Query Match             51.9%;  Score 41;  DB 1;  Length 436;
  Best Local Similarity   70.0%;  Pred. No. 23;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 AHPAQRRPWR 13
            | |:|| |||
Db    208 ATPSQRGPWR 217


RESULT 6
SHK1_HUMAN
ID   SHK1_HUMAN     STANDARD;      PRT;  2161 AA.
AC   Q9Y566; Q9NYW9;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   SH3 and multiple ankyrin repeat domains protein 1 (Shank1)
DE   (Somatostatin receptor interacting protein) (SSTR interacting protein)
DE   (SSTRIP).
GN   SHANK1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND INTERACTION WITH SSTR2.
RC   TISSUE=Fetal brain, Hippocampus, and Thalamus;
RX   MEDLINE=20020275; PubMed=10551867;
RA   Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT   "Somatostatin receptor interacting protein defines a novel family of
RT   multidomain proteins present in human and rodent brain.";
RL   J. Biol. Chem. 274:32997-33001(1999).
RN   [2]
RP   REVIEW.
RX   PubMed=10806096;
RA   Sheng M., Kim E.;
RT   "The Shank family of scaffold proteins.";
RL   J. Cell Sci. 113:1851-1856(2000).
CC   -!- FUNCTION: Seems to be a an adapter protein in the postsynaptic
CC       density (PSD) of excitatory synapses that interconnects receptors
CC       of the postsynaptic membrane including NMDA-type and metabotropic
CC       glutamate receptors via complexes with GKAP/PSD-95 and Homer,
CC       respectively, and the actin-based cytoskeleton. May play a role in
CC       the structural and functional organization of the dendritic spine
CC       and synaptic junction.
CC   -!- SUBUNIT: May homomultimerize via its SAM domain (By similarity).
CC       Interacts with SSTR2 C-terminus via the PDZ domain. Interacts with
CC       SPTAN1, Homer-1 and DLGAP1/GKAP isoforms 1 and 2 (By similarity).
CC       Is part of a complex with DLG4/PSD-95 and DLGAP1/GKAP (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
CC       neuronal cells (By similarity).
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/a (shown here), 2/b and 3; are
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in brain particularly in the
CC       amygdala, hippocampus, substantia nigra and thalamus. Isoform 2
CC       seems to be expressed ubiquitously.
CC   -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC   -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF163302; AAD45121.1; -.
DR   EMBL; AF226728; AAF35887.1; -.
DR   HSSP; P06241; 1SHF.
DR   MIM; 604999; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR001478; PDZ.
DR   InterPro; IPR001660; SAM.
```

```
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00023; ank; 7.
DR   Pfam; PF00595; PDZ; 1.
DR   Pfam; PF00536; SAM; 1.
DR   Pfam; PF00018; SH3; 1.
DR   ProDom; PD000066; SH3; 1.
DR   SMART; SM00248; ANK; 3.
DR   SMART; SM00228; PDZ; 1.
DR   SMART; SM00454; SAM; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50106; PDZ; 1.
DR   PROSITE; PS50105; SAM_DOMAIN; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   ANK repeat; SH3 domain; Repeat; Alternative splicing.
FT   REPEAT      212    245       ANK 1.
FT   REPEAT      246    278       ANK 2.
FT   REPEAT      279    312       ANK 3.
FT   REPEAT      313    345       ANK 4.
FT   REPEAT      346    378       ANK 5.
FT   REPEAT      379    395       ANK 6.
FT   DOMAIN      554    613       SH3.
FT   DOMAIN      663    757       PDZ.
FT   DOMAIN     2098   2161       SAM.
FT   DOMAIN     1002   1007       POLY-HIS.
FT   DOMAIN     1014   1019       POLY-HIS.
FT   DOMAIN     1189   1195       POLY-GLY.
FT   DOMAIN     1709   1717       POLY-GLY.
FT   DOMAIN     1844   1854       POLY-PRO.
FT   DOMAIN     1896   1902       POLY-GLY.
FT   DOMAIN     1970   1979       POLY-SER.
FT   VARSPLIC      1    613       MISSING (IN ISOFORM 2).
FT   VARSPLIC    614    654       RSQESKQESRSDKAKRLFRHYTVGSYDSFDAPSLMDGIGPG
FT                                -> MQLMALEQRFGSGLPGGGQPLCLMMSSPLPPPPPHFSC
FT                                LPA (IN ISOFORM 2).
FT   VARSPLIC    646    654       MISSING (IN ISOFORM 3).
SQ   SEQUENCE   2161 AA;  225019 MW;  5FEFC969CBE98701 CRC64;

  Query Match             51.9%;  Score 41;  DB 1;  Length 2161;
  Best Local Similarity   50.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 AAHPAQRRPWRA 14
           | ||   :|||:
Db    1468 APHPGVSKPWRS 1479
```

```
RESULT 7
THYL_HUMAN
ID   THYL_HUMAN     STANDARD;      PRT;   242 AA.
AC   P20396;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thyroliberin precursor (Thyrotropin releasing hormone) (TRH)
DE   (Protirelin).
GN   TRH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91125361; PubMed=2126343;
RA   Yamada M., Radovick S., Wondisford F.E., Nakayama Y., Weintraub B.D.,
RA   Wilber J.F.;
RT   "Cloning and structure of human genomic DNA and hypothalamic cDNA
RT   encoding human prepro thyrotropin-releasing hormone.";
RL   Mol. Endocrinol. 4:551-556(1990).
CC   -!- FUNCTION: TRH FUNCTIONS AS A REGULATOR OF THE BIOSYNTHESIS OF TSH
CC       IN THE ANTERIOR PITUITARY GLAND AND AS A NEUROTRANSMITTER/
CC       NEUROMODULATOR IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS.
CC   -!- TISSUE SPECIFICITY: HYPOTHALAMUS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63582; AAA36480.1; -.
DR   EMBL; M63581; AAA36480.1; JOINED.
DR   PIR; A34550; A34550.
DR   Genew; HGNC:12298; TRH.
DR   MIM; 275120; -.
KW   Amidation; Hormone; Repeat; Hypothalamus; Signal;
KW   Cleavage on pair of basic residues.
FT   SIGNAL        1    ?24       POTENTIAL.
FT   CHAIN        25    242       PROTHYROLIBERIN.
FT   PEPTIDE      84     86       THYROLIBERIN.
FT   PEPTIDE     114    116       THYROLIBERIN.
FT   PEPTIDE     135    137       THYROLIBERIN.
FT   PEPTIDE     152    154       THYROLIBERIN.
FT   PEPTIDE     186    188       THYROLIBERIN.
FT   PEPTIDE     227    229       THYROLIBERIN.
FT   MOD_RES      84     84       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      86     86       AMIDATION (G-87 PROVIDE AMIDE GROUP).
FT   MOD_RES     114    114       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES     116    116       AMIDATION (G-117 PROVIDE AMIDE GROUP).
FT   MOD_RES     135    135       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES     137    137       AMIDATION (G-138 PROVIDE AMIDE GROUP).
FT   MOD_RES     152    152       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES     154    154       AMIDATION (G-155 PROVIDE AMIDE GROUP).
FT   MOD_RES     186    186       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES     188    188       AMIDATION (G-189 PROVIDE AMIDE GROUP).
FT   MOD_RES     227    227       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES     229    229       AMIDATION (G-230 PROVIDE AMIDE GROUP).
SQ   SEQUENCE   242 AA;  27404 MW;  8C0F9D915B32F29F CRC64;

  Query Match             50.6%;  Score 40;  DB 1;  Length 242;
  Best Local Similarity   60.0%;  Pred. No. 19;
  Matches    6;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 HPAQRRPWRA 14
           || :| || |
Db     136 HPGRRSPWLA 145
```

```
RESULT 8
Y093_RHIME
ID   Y093_RHIME     STANDARD;      PRT;   465 AA.
AC   O87394;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical transport protein R00093.
GN   R00093 OR SMC02616.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RA   Powers E.L., Vuyyuru V., Kahn M.L.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1021;
RX   MEDLINE=21396507; PubMed=11481430;
RA   Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA   Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA   Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
```

```
RA   Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA   Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT   "Analysis of the chromosome sequence of the legume symbiont
RT   Sinorhizobium meliloti strain 1021.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC   -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF055582; AAC62224.1; -.
DR   EMBL; AL591782; CAC41480.1; -.
DR   InterPro; IPR002293; AA/rel_prmease1.
DR   InterPro; IPR004840; AAc_permease.
DR   InterPro; IPR004841; Permease.
DR   Pfam; PF00324; aa_permeases; 1.
DR   PROSITE; PS00218; AMINO_ACID_PERMEASE_1; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT   TRANSMEM     19     39       POTENTIAL.
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM     91    111       POTENTIAL.
FT   TRANSMEM    140    160       POTENTIAL.
FT   TRANSMEM    164    184       POTENTIAL.
FT   TRANSMEM    201    221       POTENTIAL.
FT   TRANSMEM    244    264       POTENTIAL.
FT   TRANSMEM    288    308       POTENTIAL.
FT   TRANSMEM    342    362       POTENTIAL.
FT   TRANSMEM    363    383       POTENTIAL.
FT   TRANSMEM    403    423       POTENTIAL.
SQ   SEQUENCE   465 AA;  50783 MW;  0375E164F737AA0A CRC64;

  Query Match           50.6%;  Score 40;  DB 1;  Length 465;
  Best Local Similarity   58.3%;  Pred. No. 35;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 AAHPAQRRPWRA 14
           | ||  | |:||
Db     330 AVHPVYRTPYRA 341


RESULT 9
HMW3_MYCGE
ID   HMW3_MYCGE     STANDARD;      PRT;   599 AA.
AC   Q57081; Q49337; Q49191; Q49370;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytadherence high molecular weight protein 3 (Cytadherence accessory
DE   protein 3) (Accessory adhesin protein 3) (P69).
GN   HMW3 OR MG317.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96011386; PubMed=7592348;
RA   Reddy S.P., Rasmussen W.G., Baseman J.B.;
RT   "Molecular cloning and characterization of an adherence-related
RT   operon of Mycoplasma genitalium.";
RL   J. Bacteriol. 177:5943-5951(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [3]
RP   SEQUENCE OF 1-24; 57-169 AND 444-514 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC       STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC       CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC       PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC       IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE.
CC       ESSENTIAL FOR SUCCESFUL SURFACE PARASITISM (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC       MEMBRANE (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39712; AAC71539.1; -.
DR   EMBL; L43097; AAA99946.1; -.
DR   EMBL; U01716; AAC43190.1; ALT_INIT.
DR   EMBL; U02224; AAA03378.1; -.
DR   EMBL; U02267; AAD12533.1; -.
DR   TIGR; MG317; -.
KW   Cytadherence; Structural protein; Complete proteome.
SQ   SEQUENCE   599 AA;  68720 MW;  D786BE7BD491129A CRC64;

  Query Match           50.6%;  Score 40;  DB 1;  Length 599;
  Best Local Similarity   66.7%;  Pred. No. 44;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 HPAQRRPWR 13
           :|  |||||
Db     544 YPLTRRPWR 552


RESULT 10
YPU3_RHOCA
ID   YPU3_RHOCA     STANDARD;      PRT;    55 AA.
AC   P26159;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 5.8 kDa protein in PUHA 5'region (ORF55).
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Burke D.H., Alberti M., Armstrong G.A., Hearst J.E.;
RL   Submitted (NOV-1991) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; Z11165; CAA77517.1; -.
DR   PIR; S17805; S17805.
KW   Photosynthesis; Hypothetical protein.
SQ   SEQUENCE   55 AA;  5750 MW;  7EB55296266D48B1 CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 55;
  Best Local Similarity   60.0%;  Pred. No. 6.5;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 AAHPAQRRPW 12
          :| ||:|:||
Db     12 SAAPARRQPW 21


RESULT 11
DRAG_RHORU
ID   DRAG_RHORU     STANDARD;      PRT;   294 AA.
AC   P14300;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ADP-ribosyl-[dinitrogen reductase] glycohydrolase (EC 3.2.2.24) (ADP-
DE   ribosylglycohydrolase) (Dinitrogenase reductase activating
DE   glycohydrolase).
GN   DRAG.
OS   Rhodospirillum rubrum.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC   Rhodospirillum.
OX   NCBI_TaxID=1085;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 3-35 AND 36-38.
RC   STRAIN=UR2;
RX   MEDLINE=89384461; PubMed=2506427;
RA   Fitzmaurice W.P., Saari L.L., Lowery R.G., Ludden P.W., Roberts G.P.;
RT   "Genes coding for the reversible ADP-ribosylation system of
RT   dinitrogenase reductase from Rhodospirillum rubrum.";
RL   Mol. Gen. Genet. 218:340-347(1989).
CC   -!- FUNCTION: Involved in the regulation of the nitrogen fixation
CC       activity by the reversible ADP-ribosylation of the dinitrogenase
CC       reductase component of the nitrogenase enzyme complex. The ADP-
CC       ribosyltransferase (DraT) transfers the ADP-ribose group from NAD
CC       to dinitrogenase reductase. The ADP-ribose group is removed
CC       through the action of the ADP-ribosylglycohydrolase (DraG).
CC   -!- CATALYTIC ACTIVITY: ADP-D-ribosyl-[dinitrogen reductase] =
CC       [dinitrogen reductase] + ADP-ribose.
CC   -!- SIMILARITY: BELONGS TO THE DRAG FAMILY.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; X16187; CAA34310.1; -.
DR   PIR; JT0536; JT0536.
KW   Hydrolase; Nitrogen fixation.
SQ   SEQUENCE   294 AA;  31792 MW;  5E72ECFA8A798368 CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 294;
  Best Local Similarity   54.5%;  Pred. No. 32;
  Matches    6;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 TAAHPAQRRPW 12
          |  |||   ||
Db    139 TLGHPADLEPW 149


RESULT 12
PENA_BURCE
ID   PENA_BURCE     STANDARD;      PRT;   313 AA.
AC   Q02940;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Beta-lactamase precursor (EC 3.5.2.6) (Penicillinase).
GN   PENA.
OS   Burkholderia cepacia (Pseudomonas cepacia).
OC   Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC   Burkholderia.
OX   NCBI_TaxID=292;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 17616 / 249;
RX   MEDLINE=93263630; PubMed=8494361;
RA   Proenca R., Niu W.W., Cacalano G., Prince A.;
RT   "The Pseudomonas cepacia 249 chromosomal penicillinase is a member of
RT   the AmpC family of chromosomal beta-lactamases.";
RL   Antimicrob. Agents Chemother. 37:667-674(1993).
CC   -!- FUNCTION: ENABLES THE ORGANISM TO UTILIZE PENICILLIN AS A
CC       CARBON SOURCE.
CC   -!- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
CC       amino acid.
CC   -!- INDUCTION: BY PENICILLIN G, IMPENEM AND AMPR.
CC   -!- SIMILARITY: BELONGS TO THE CLASS-C BETA-LACTAMASE FAMILY.
CC   -!- CAUTION: THIS PROTEIN COULD BE ARTIFACTUAL, IT SEEMS TO CONTAIN
CC       PIECES OD SEVERAL DIFFERENT PROTEINS.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; L02928; AAA25927.1; -.
DR   PIR; A48903; A48903.
DR   InterPro; IPR001586; Beta_lactamase_C.
DR   PROSITE; PS00336; BETA_LACTAMASE_C; 1.
KW   Antibiotic resistance; Hydrolase; Signal.
FT   SIGNAL        1     15
FT   CHAIN        16    313       BETA-LACTAMASE.
FT   ACT_SITE    190    190       POTENTIAL.
SQ   SEQUENCE   313 AA;  34327 MW;  BF2A67C670A644F2 CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 313;
  Best Local Similarity   58.3%;  Pred. No. 34;
  Matches    7;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 RTAAHPAQRRPW 12
          |  | ||:| ||
Db    243 RGPAEPAKRGPW 254


RESULT 13
APBE_CHLPN
ID   APBE_CHLPN     STANDARD;      PRT;   314 AA.
AC   Q9Z8K2; Q9JQA3;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thiamine biosynthesis lipoprotein apbE precursor.
GN   APBE OR CPN0336 OR CP0422.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
```

```
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- FUNCTION: INVOLVED IN THE CONVERSION OF AMINOIMIDAZOLE RIBOTIDE
CC       (AIR), A PURINE INTERMEDIATE, TO THE 4-AMINO-5-HYDROXYMETHYL-2-
CC       METHYL PYRIMIDINE (HMP) MOIETY OF THIAMINE (BY SIMILARITY).
CC   -!- PATHWAY: BIOSYNTHESIS OF THE PYRIMIDINE MOIETY OF THIAMINE.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE APBE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001618; AAD18485.1; -.
DR   EMBL; AE002203; AAF38265.1; -.
DR   EMBL; AP002546; BAA98546.1; -.
DR   TIGR; CP0422; -.
DR   InterPro; IPR003374; ApbE.
DR   Pfam; PF02424; ApbE; 1.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Thiamine biosynthesis; Membrane; Lipoprotein; Signal;
KW   Complete proteome.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    314       THIAMINE BIOSYNTHESIS LIPOPROTEIN APBE.
FT   LIPID        19     19       N-ACYL DIGLYCERIDE (POTENTIAL).
SQ   SEQUENCE   314 AA;  35356 MW;  3549A571FC475FCF CRC64;

  Query Match             49.4%;  Score 39;  DB 1;  Length 314;
  Best Local Similarity   46.2%;  Pred. No. 34;
  Matches    6;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWR 13
           :|: |    ||||
Db     201 KTSGHHPSGRPWR 213


RESULT 14
NTC4_MOUSE
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; Q62389; Q62390; O35442; Q9R1W9; O88314; O88316; Q9R1X0;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 4 precursor (Notch 4)
DE   [Contains: Transforming protein Int-3].
GN   NOTCH4 OR INT3 OR INT-3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92194507; PubMed=1312643;
RA   Robbins J., Blondel B.J., Gallahan D., Callahan R.;
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
RL   J. Virol. 66:2594-2599(1992).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RX   MEDLINE=97294599; PubMed=9150355;
RA   Gallahan D., Callahan R.;
RT   "The mouse mammary tumor associated gene INT3 is a unique member of
RT   the NOTCH gene family (NOTCH4).";
RL   Oncogene 14:1883-1890(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung, and Testis;
RX   MEDLINE=96281668; PubMed=8681805;
RA   Uyttendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT   "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT   cell-specific mammalian Notch gene.";
RL   Development 122:2251-2259(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA   Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RT   "Sequence of the mouse major histocompatibility locus class III
RT   region.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1436-1600 FROM N.A.
RX   MEDLINE=99252212; PubMed=10233982;
RA   Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
RT   "Intracisternal type A particle-mediated activation of the Notch4/int3
RT   gene in a mouse mammary tumor: generation of truncated Notch4/int3
RT   mRNAs by retroviral splicing events.";
RL   J. Virol. 73:5166-5171(1999).
RN   [6]
RP   FUNCTION.
RX   MEDLINE=21244657; PubMed=11344305;
RA   Uyttendaele H., Ho J., Rossant J., Kitajewski J.;
RT   "Vascular patterning defects associated with expression of activated
RT   Notch4 in embryonic endothelium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
RN   [7]
RP   SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
RP   OF VAL-1463.
RX   MEDLINE=21523956; PubMed=11518718;
RA   Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT   "Murine notch homologs (N1-4) undergo presenilin-dependent
RT   proteolysis.";
RL   J. Biol. Chem. 276:40268-40273(2001).
RN   [8]
RP   POST-TRANSLATIONAL PROCESSING.
RX   MEDLINE=21374376; PubMed=11459941;
RA   Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT   "Conservation of the biochemical mechanisms of signal transduction
RT   among mammalian Notch family members.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). May regulate branching
CC       morphogenesis in the developing vascular system.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds.
```

```
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
CC       kidney, and at lower levels in the ovary and skeletal muscle. A
CC       very low expression is seen in the brain, intestine, liver and
CC       testis.
CC   -!- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
CC       embryonic development from 9.0 d.p.c.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane.
CC   -!- PTM: Phosphorylated.
CC   -!- DISEASE: Loss of the extracellular domain causes constitutive
CC       activation of the Notch protein, which leads to hyperproliferation
CC       of glandular epithelial tissues and development of mammary
CC       carcinomas.
CC   -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC   -!- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80456; AAB38377.1; -.
DR   EMBL; U43691; AAC52630.1; -.
DR   EMBL; U43691; AAC52631.1; -.
DR   EMBL; AF030001; AAB82004.1; -.
DR   EMBL; AB016771; BAA32281.1; ALT_SEQ.
DR   EMBL; AB016772; BAA32283.1; ALT_INIT.
DR   EMBL; AB016773; BAA32284.1; ALT_INIT.
DR   EMBL; AB016774; BAA32285.1; -.
DR   PIR; A38072; TVMVT3.
DR   HSSP; P08709; 1BF9.
DR   MGD; MGI:107471; Notch4.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000800; Notch.
DR   Pfam; PF00008; EGF; 27.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00066; notch; 2.
DR   PRINTS; PR01415; ANKYRIN.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 5.
DR   SMART; SM00179; EGF_CA; 11.
DR   SMART; SM00001; EGF_like; 15.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1964       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT   CHAIN      1411   1964       TRANSFORMING PROTEIN INT-3.
FT   CHAIN      1428   1964       NOTCH EXTRACELLULAR TRUNCATION.
FT   CHAIN      1463   1964       NOTCH INTRACELLULAR DOMAIN.
FT   DOMAIN       21   1443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1444   1464       POTENTIAL.
FT   DOMAIN     1465   1964       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       21     60       EGF-LIKE 1.
FT   DOMAIN       61    112       EGF-LIKE 2.
FT   DOMAIN      115    152       EGF-LIKE 3.
FT   DOMAIN      153    189       EGF-LIKE 4.
FT   DOMAIN      191    229       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      231    271       EGF-LIKE 6.
FT   DOMAIN      273    309       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      311    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      352    388       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      389    427       EGF-LIKE 10.
FT   DOMAIN      429    470       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      472    508       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      510    546       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      548    584       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      586    622       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      623    656       EGF-LIKE 16.
FT   DOMAIN      658    686       EGF-LIKE 17.
FT   DOMAIN      688    724       EGF-LIKE 18.
FT   DOMAIN      726    762       EGF-LIKE 19.
FT   DOMAIN      764    800       EGF-LIKE 20.
FT   DOMAIN      803    839       EGF-LIKE 21.
FT   DOMAIN      841    877       EGF-LIKE 22.
FT   DOMAIN      878    924       EGF-LIKE 23.
FT   DOMAIN      926    962       EGF-LIKE 24.
FT   DOMAIN      964   1000       EGF-LIKE 25.
FT   DOMAIN     1002   1040       EGF-LIKE 26.
FT   DOMAIN     1042   1081       EGF-LIKE 27.
FT   DOMAIN     1083   1122       EGF-LIKE 28.
FT   DOMAIN     1126   1167       EGF-LIKE 29.
FT   REPEAT     1168   1208       LIN/NOTCH 1.
FT   REPEAT     1209   1242       LIN/NOTCH 2.
FT   REPEAT     1243   1282       LIN/NOTCH 3.
FT   REPEAT     1628   1657       ANK 1.
FT   REPEAT     1661   1691       ANK 2.
FT   REPEAT     1695   1724       ANK 3.
FT   REPEAT     1728   1757       ANK 4.

  Query Match             49.4%;  Score 39;  DB 1;  Length 1964;
  Best Local Similarity   75.0%;  Pred. No. 2e+02;
  Matches    6;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 HPAQRRPW 12
           ||| | ||
Db    1392 HPASRCPW 1399


RESULT 15
NTC4_HUMAN
ID   NTC4_HUMAN      STANDARD;      PRT;  2003 AA.
AC   Q99466; O00306; Q99940; Q99458; Q9H3S8; Q9UII9; Q9UIJ0;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 4 precursor (Notch 4)
DE   (hNotch4).
GN   NOTCH4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
RC   TISSUE=Placenta;
RX   MEDLINE=97311416; PubMed=9168133;
```

```
RA   Sugaya K., Sasanuma S.-I., Nohata J., Kimura T., Fukagawa T.,
RA   Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
RT   "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
RT   human counterpart gene of mouse proto-oncogene Int3.";
RL   Gene 189:235-244(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Bone marrow, and Heart;
RX   MEDLINE=98360091; PubMed=9693032;
RA   Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
RA   Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
RT   "Cloning, characterization, and the complete 56.8-kilobase DNA
RT   sequence of the human NOTCH4 gene.";
RL   Genomics 51:45-58(1998).
RN   [3]
RP   SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
RA   Miyagawa T., Tokunaga K., Hojho H.;
RT   "Human notch4 gene variant.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   IDENTIFICATION OF LIGANDS.
RX   MEDLINE=99180765; PubMed=10079256;
RA   Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA   Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT   "Human ligands of the Notch receptor.";
RL   Am. J. Pathol. 154:785-794(1999).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs. May regulate branching morphogenesis in the
CC       developing vascular system (By similarity).
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; may be
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
CC       the lung and placenta and at low levels in the liver, skeletal
CC       muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
CC       and fetal liver. No expression was seen in adult brain or
CC       peripheral blood leukocytes.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
CC       peptide) is polymorphic and the number of Leu varies in the
CC       population (from 6 to 12).
CC   -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC   -!- SIMILARITY: CONTAINS 28 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC       in position 1438 to 1463.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D63395; BAA09708.1; ALT_FRAME.
DR   EMBL; D86566; BAA13116.1; -.
DR   EMBL; U95299; AAC32288.1; -.
DR   EMBL; U89335; AAC63097.1; -.
DR   EMBL; AB023961; BAB20317.1; -.
DR   EMBL; AB024520; BAA88951.1; -.
DR   EMBL; AB024578; BAA88952.1; -.
DR   HSSP; P08709; 1BF9.
DR   Genew; HGNC:7884; NOTCH4.
DR   MIM; 164951; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000800; Notch.
DR   Pfam; PF00008; EGF; 26.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00066; notch; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00012; FNTYPEI.
DR   SMART; SM00248; ANK; 5.
DR   SMART; SM00179; EGF_CA; 11.
DR   SMART; SM00001; EGF_like; 15.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS01187; EGF_CA; 9.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW   Triplet repeat expansion; Alternative splicing.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   2003       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
FT   CHAIN      1432   2003       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1467   2003       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       24   1447       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1448   1468       POTENTIAL.
FT   DOMAIN     1469   2003       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       24     63       EGF-LIKE 1.
FT   DOMAIN       64    115       EGF-LIKE 2.
FT   DOMAIN      118    155       EGF-LIKE 3.
FT   DOMAIN      156    192       EGF-LIKE 4.
FT   DOMAIN      194    232       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      234    274       EGF-LIKE 6.
FT   DOMAIN      276    312       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      314    353       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      355    391       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      392    430       EGF-LIKE 10.
FT   DOMAIN      432    473       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      475    511       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      513    549       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      551    587       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      589    625       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      626    659       EGF-LIKE 16.
FT   DOMAIN      661    689       EGF-LIKE 17.
FT   DOMAIN      691    727       EGF-LIKE 18.
FT   DOMAIN      729    765       EGF-LIKE 19.
FT   DOMAIN      767    803       EGF-LIKE 20.
FT   DOMAIN      806    842       EGF-LIKE 21.
FT   DOMAIN      844    880       EGF-LIKE 22.
FT   DOMAIN      882    928       EGF-LIKE 23.
FT   DOMAIN      930    966       EGF-LIKE 24.
```

FT	DOMAIN	968	1004	EGF-LIKE 25.
FT	DOMAIN	1006	1044	EGF-LIKE 26.
FT	DOMAIN	1046	1085	EGF-LIKE 27.
FT	DOMAIN	1087	1126	EGF-LIKE 28.
FT	DOMAIN	1130	1171	EGF-LIKE 29.
FT	DOMAIN	1472	1476	POLY-ARG.
FT	REPEAT	1165	1212	LIN/NOTCH 1.
FT	REPEAT	1213	1246	LIN/NOTCH 2.
FT	REPEAT	1247	1286	LIN/NOTCH 3.
FT	REPEAT	1633	1665	ANK 1.
FT	REPEAT	1666	1698	ANK 2.
FT	REPEAT	1700	1732	ANK 3.
FT	REPEAT	1733	1765	ANK 4.
FT	REPEAT	1766	1798	ANK 5.
FT	DISULFID	28	41	BY SIMILARITY.
FT	DISULFID	35	51	BY SIMILARITY.
FT	DISULFID	53	62	BY SIMILARITY.
FT	DISULFID	68	80	BY SIMILARITY.
FT	DISULFID	74	103	BY SIMILARITY.
FT	DISULFID	105	114	BY SIMILARITY.
FT	DISULFID	122	133	BY SIMILARITY.
FT	DISULFID	127	143	BY SIMILARITY.
FT	DISULFID	145	154	BY SIMILARITY.
FT	DISULFID	160	171	BY SIMILARITY.
FT	DISULFID	165	180	BY SIMILARITY.
FT	DISULFID	182	191	BY SIMILARITY.
FT	DISULFID	198	211	BY SIMILARITY.
FT	DISULFID	205	220	BY SIMILARITY.
FT	DISULFID	222	231	BY SIMILARITY.
FT	DISULFID	238	249	BY SIMILARITY.
FT	DISULFID	243	262	BY SIMILARITY.
FT	DISULFID	264	273	BY SIMILARITY.
FT	DISULFID	280	291	BY SIMILARITY.
FT	DISULFID	285	300	BY SIMILARITY.
FT	DISULFID	302	311	BY SIMILARITY.
FT	DISULFID	318	332	BY SIMILARITY.
FT	DISULFID	326	341	BY SIMILARITY.
FT	DISULFID	343	352	BY SIMILARITY.

Query Match 49.4%; Score 39; DB 1; Length 2003;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       5 HPAQRRPW 12
           ||| | ||
Db    1396 HPASRCPW 1403
```

[redacted]arch completed: March 10, 2003, 17:13:52
[redacted]b time : 10.6154 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: March 10, 2003, 16:57:56 ; Search time 32.6667 Seconds
(without alignments)
57.107 Million cell updates/sec

Title: US-09-993-392-3
Perfect score: 79
Sequence: 1 RTAAHPAQRRPWRA 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	50	63.3	55	22	AAU60174	Propionibacterium
2	48	60.8	192	22	AAU25578	Human G Protein-Co
3	46	58.2	359	22	ABB30654	Peptide #3305 enco
4	46	58.2	359	22	ABB35825	Peptide #3331 enco
5	46	58.2	359	22	ABB21241	Protein #3240 enco
6	46	58.2	359	22	AAM56631	Human brain expres
7	46	58.2	359	22	AAM69009	Human bone marrow
8	46	58.2	359	22	AAM16842	Peptide #3276 enco
9	46	58.2	359	22	AAM29327	Peptide #3364 enco
10	46	58.2	359	22	AAM04552	Peptide #3234 enco
11	46	58.2	359	23	ABG38604	Human peptide enco
12	45	57.0	405	22	ABG12849	Novel human diagno
13	45	57.0	903	6	AAP50312	Herpes simplex vir
14	44	55.7	49	22	AAU29989	Novel human secret
15	44	55.7	59	22	AAU53809	Propionibacterium
16	44	55.7	1250	22	ABB12254	Human S3-12 homolo
17	43	54.4	64	22	AAU62378	Propionibacterium
18	43	54.4	365	22	AAU50463	Propionibacterium
19	42	53.2	253	23	AAG80770	L. esculentum expa
20	42	53.2	255	22	ABB65870	Drosophila melanog
21	42	53.2	377	22	AAB96367	Putative P. abyssi
22	42	53.2	393	20	AAY39471	Maize uroporphyrin
23	41.5	52.5	203	22	AAU31851	Novel human secret
24	41	51.9	182	22	AAM42034	Human polypeptide
25	41	51.9	224	22	AAU61782	Propionibacterium
26	41	51.9	230	22	AAB63265	Human breast cance
27	41	51.9	274	22	AAU01012	PEPc kinase fragme
28	41	51.9	487	22	AAE05091	Rice SPF1-related
29	41	51.9	488	22	AAE05090	Rice SPF1-related
30	41	51.9	569	22	ABG29400	Novel human diagno
31	41	51.9	635	18	AAW19920	Human Ksr' (kinase
32	41	51.9	641	22	ABB11444	Human Ksr-1 homolo
33	41	51.9	641	22	AAM80115	Human protein SEQ
34	41	51.9	764	22	AAM79131	Human protein SEQ
35	41	51.9	799	22	AAU49462	Propionibacterium
36	41	51.9	873	18	AAW19918	Mouse Ksr-1 (kinas
37	41	51.9	875	18	AAW19919	Human Ksr-1 (kinas
38	40.5	51.3	60	22	AAU49983	Propionibacterium
39	40	50.6	60	21	AAY65120	Human 5' EST relat
40	40	50.6	88	22	AAU50493	Propionibacterium
41	40	50.6	93	22	AAU52758	Propionibacterium
42	40	50.6	104	22	AAG77462	Human colon cancer
43	40	50.6	106	22	AAU39141	Propionibacterium
44	40	50.6	118	23	ABP38025	Staphylococcus epi
45	40	50.6	141	22	ABG20345	Novel human diagno

ALIGNMENTS

RESULT 1
AAU60174
ID AAU60174 standard; Protein; 55 AA.
XX
AC AAU60174;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #21070.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.

```
DR   N-PSDB; AAS59608.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 21369; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   55 AA;

  Query Match             63.3%;  Score 50;  DB 22;  Length 55;
  Best Local Similarity   75.0%;  Pred. No. 1.5;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 AAHPAQRRPWRA 14
          |  || ||||||
Db     21 AGQPAVRRPWRA 32


RESULT 2
AAU25578
ID   AAU25578 standard; Protein; 192 AA.
XX
AC   AAU25578;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human G Protein-Coupled Receptor (GPCR) polypeptide #25.
XX
KW   Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;
KW   attention deficit disorder; anxiety; depression; bipolar disorder;
KW   neurological disorder; Huntington's disease; dementia; obesity; anorexia;
KW   metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;
KW   type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;
KW   cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;
KW   viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;
KW   antidepressant; anorectic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200162797-A2.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-US05676.
XX
PR   23-FEB-2000; 2000US-0184247.
PR   23-FEB-2000; 2000US-0184303.
PR   23-FEB-2000; 2000US-0184304.
PR   23-FEB-2000; 2000US-0184305.
PR   23-FEB-2000; 2000US-0184397.
PR   02-MAR-2000; 2000US-0186457.
PR   03-MAR-2000; 2000US-0186810.
PR   09-MAR-2000; 2000US-0188064.
PR   13-MAR-2000; 2000US-0188880.
PR   03-APR-2000; 2000US-0194344.
PR   23-JUN-2000; 2000US-0213861.
PR   11-JUL-2000; 2000US-0217369.
PR   11-JUL-2000; 2000US-0217370.
PR   14-JUL-2000; 2000US-0218337.
PR   20-JUL-2000; 2000US-0218492.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Vogeli G,  Wood LS,  Parodi LA,  Lind P;
XX
DR   WPI; 2001-570628/64.
DR   N-PSDB; AAS42830.
XX
PT   New isolated nucleic acid encoding a new G-protein coupled receptor
PT   polypeptide for detecting receptor modulators that can treat mental
PT   disorders, such as schizophrenia, anxiety, depression, or obesity -
XX
PS   Claim 35; Page 79; 279pp; English.
XX
CC   Sequences AAU25554-AAU25616 represent human G-protein coupled receptor
CC   (GPCR) polypeptides of the invention. The proteins and their associated
CC   DNA sequences can be used to identify compounds which bind to GPCR
CC   polypeptides and in screening for compounds that modulate GPCR activity.
CC   By screening a human subject for the presence of mutations in GPCR DNA, a
CC   GPCR-related disorder or a genetic predisposition can be diagnosed. The
CC   sequences can also be used for treatment and prevention of mental
CC   disorders such as schizophrenia, attention deficit disorder, anxiety,
CC   depression, dementia and bipolar disorder, neurological disorders such as
CC   Huntington's disease, Parkinson's disease and Tourette's syndrome,
CC   metabolic disorders such as obesity, anorexia and type 2 diabetes,
CC   cardiovascular disorders such as thrombosis, myocardial infarction,
CC   cardiomyopathy and atherosclerosis, viral infections caused by HIV and
CC   cancers.
XX
SQ   Sequence   192 AA;

  Query Match             60.8%;  Score 48;  DB 22;  Length 192;
  Best Local Similarity   63.6%;  Pred. No. 10;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 TAAHPAQRRPW 12
          |: ||  ||||
Db     52 TSTHPLSRRPW 62


RESULT 3
ABB30654
ID   ABB30654 standard; Peptide; 359 AA.
XX
AC   ABB30654;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Peptide #3305 encoded by breast cell single exon nucleic acid probe.
XX
KW   Human; microarray; single exon probe; gene expression; breast;
KW   disease; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157271-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00662.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
```

PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 27; SEQ ID NO 13622; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 359 AA;

Query Match 58.2%; Score 46; DB 22; Length 359;
Best Local Similarity 72.7%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAHPAQRRPWR 13
||||| | ||
Db 147 AAHPAHSRAWR 157

RESULT 4
ABB35825
ID ABB35825 standard; Peptide; 359 AA.
XX
AC ABB35825;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #3331 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 28460; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 359 AA;

Query Match 58.2%; Score 46; DB 22; Length 359;
Best Local Similarity 72.7%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAHPAQRRPWR 13
||||| | ||
Db 147 AAHPAHSRAWR 157

RESULT 5
ABB21241
ID ABB21241 standard; Protein; 359 AA.
XX
AC ABB21241;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #3240 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -

XX
PS Claim 15; SEQ ID No 23011; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 359 AA;

Query Match 58.2%; Score 46; DB 22; Length 359;
Best Local Similarity 72.7%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAHPAQRRPWR 13
||||| | ||
Db 147 AAHPAHSRAWR 157

RESULT 6
AAM56631
ID AAM56631 standard; Protein; 359 AA.
XX
AC AAM56631;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 28736.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 28736; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 359 AA;

Query Match 58.2%; Score 46; DB 22; Length 359;
Best Local Similarity 72.7%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAHPAQRRPWR 13
||||| | ||
Db 147 AAHPAHSRAWR 157

RESULT 7
AAM69009
ID AAM69009 standard; Protein; 359 AA.
XX
AC AAM69009;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29315.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 29315; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 359 AA;

Query Match 58.2%; Score 46; DB 22; Length 359;
Best Local Similarity 72.7%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AAHPAQRRPWR 13
||||| | ||
Db 147 AAHPAHSRAWR 157

RESULT 8

```
AAM16842
ID   AAM16842 standard; Protein; 359 AA.
XX
AC   AAM16842;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #3276 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 21668; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   359 AA;

  Query Match             58.2%;  Score 46;  DB 22;  Length 359;
  Best Local Similarity   72.7%;  Pred. No. 38;
  Matches    8;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 AAHPAQRRPWR 13
           |||||  | ||
Db     147 AAHPAHSRAWR 157


RESULT 9
AAM29327
ID   AAM29327 standard; Protein; 359 AA.
XX
AC   AAM29327;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Peptide #3364 encoded by probe for measuring placental gene expression.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 27; SEQ ID No 29596; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP:
CC   see AAI31315-AAI57546). The present sequence is a peptide encoded by one
CC   such probe. The probes are useful for producing a microarray for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from human placenta. The probes are useful for antenatal diagnosis of
CC   human genetic disorders.
XX
SQ   Sequence   359 AA;

  Query Match             58.2%;  Score 46;  DB 22;  Length 359;
  Best Local Similarity   72.7%;  Pred. No. 38;
  Matches    8;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 AAHPAQRRPWR 13
           |||||  | ||
Db     147 AAHPAHSRAWR 157


RESULT 10
AAM04552
ID   AAM04552 standard; Protein; 359 AA.
XX
AC   AAM04552;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Peptide #3234 encoded by probe for measuring breast gene expression.
XX
KW   Probe; human; breast disease; breast cancer; development disorder;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 27; SEQ ID No 13292; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes
CC (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 359 AA;

Query Match 58.2%; Score 46; DB 22; Length 359;
Best Local Similarity 72.7%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      3 AAHPAQRRPWR 13
          |||||  | ||
Db    147 AAHPAHSRAWR 157
```

RESULT 11
ABG38604
ID ABG38604 standard; Peptide; 359 AA.
XX
AC ABG38604;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human peptide encoded by genome-derived single exon probe SEQ ID 28269.
XX
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomtosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesis; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 27; SEQ ID No 28269; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomtosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesis, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 359 AA;

Query Match 58.2%; Score 46; DB 23; Length 359;
Best Local Similarity 72.7%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      3 AAHPAQRRPWR 13
          |||||  | ||
Db    147 AAHPAHSRAWR 157
```

RESULT 12
ABG12849
ID ABG12849 standard; Protein; 405 AA.
XX
AC ABG12849;
XX
DT 13-FEB-2002 (first entry)
XX

```
DE   Novel human diagnostic protein #12840.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;

     WPI; 2001-639362/73.
     N-PSDB; AAS77036.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 43208; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   405 AA;

  Query Match             57.0%;  Score 45;  DB 22;  Length 405;
  Best Local Similarity   61.5%;  Pred. No. 60;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       2 TAAHPAQRRPWRA 14
           |  | | |||| |
Db     325 TGVHEAVRRPWAA 337


RESULT 13
AAP50312
ID   AAP50312 standard; protein; 903 AA.
XX
AC   AAP50312;
XX
DT   12-NOV-1991  (first entry)
XX
DE   Herpes simplex virus 1 glycoprotein B.
XX
KW   Herpes simplex virus 1; glycoprotein B; vaccine;
XX
OS   Herpes simplex virus 1.
XX
PN   WO8504587-A.
XX
PD   24-OCT-1985.
XX
PF   04-APR-1985;   85WO-US00587.
XX
PR   06-APR-1984;   84US-0597784.
PR   17-JUL-1984;   84US-0631669.
XX
PA   (CHIR-) CHIRON CORP.
XX
PI   Burke RL,  Pachl C,  Valenzuela PDT,  Urdea MS;
XX
DR   WPI; 1985-276087/44.
DR   N-PSDB; AAN50364.
XX
PT   Recombinant herpes simplex vaccine - prepd. by expression of DNA
PT   constructs in a eukaryotic host.
XX
PS   Disclosure; Table 1 page 26-30; 80pp; English.
XX
CC   Herpes simplex virus glycoprotein B or fragments may be used in a
CC   vaccine against HSV. Dosage is 10 micrograms to 2 mg/kg. The
CC   glycoprotein DNA is expressed in an eukaryotic host, esp.
CC   Saccharomyces cerevisiae, CHO cells and COS cells. Suitable plasmids
CC   are pYHS115, 116, 117, 118 and 119.
XX
SQ   Sequence   903 AA;

  Query Match             57.0%;  Score 45;  DB 6;  Length 903;
  Best Local Similarity   80.0%;  Pred. No. 1.3e+02;
  Matches    8;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 AHPAQRRPWR 13
           | || |||||
Db      52 ATPAPRRPWR 61


RESULT 14
AAU29989
ID   AAU29989 standard; Protein; 49 AA.
XX
AC   AAU29989;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #480.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
PF   16-APR-2001; 2001WO-US08656.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-611725/70.
XX
```

PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
XX
PS Claim 20; Page 217; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX
SQ Sequence 49 AA;

Query Match 55.7%; Score 44; DB 22; Length 49;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RTAAHPAQRRPWRA 14
|: || : ||| |
Db 35 RSYTHPLKARPWSA 48

RESULT 15
AAU53809
ID AAU53809 standard; Protein; 59 AA.
XX
AC AAU53809;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #14705.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59562.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 15004; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 59 AA;

Query Match 55.7%; Score 44; DB 22; Length 59;
Best Local Similarity 53.8%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RTAAHPAQRRPWR 13
| : | :|||||
Db 40 RRSDRPTRRRPWR 52

Search completed: March 10, 2003, 17:13:15
Job time : 34.6667 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: March 10, 2003, 17:01:31 ; Search time 24.7692 Seconds
(without alignments)
116.461 Million cell updates/sec

Title: US-09-993-392-3
Perfect score: 79
Sequence: 1 RTAAHPAQRRPWRA 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

al number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	53	67.1	367	10	Q9SM14	Q9sm14 zea mays (m
2	46	58.2	454	13	Q90Z68	Q90z68 brachydanio
3	45	57.0	358	11	Q9CRM1	Q9crm1 mus musculu
4	45	57.0	949	11	Q8R2G6	Q8r2g6 mus musculu
5	44	55.7	106	12	O92501	O92501 bombyx mori
6	44	55.7	259	16	Q9ABC2	Q9abc2 caulobacter
7	44	55.7	363	16	Q9RV04	Q9rv04 deinococcus
8	44	55.7	2160	5	Q9U1K8	Q9u1k8 drosophila
9	43.5	55.1	326	16	Q9L2G2	Q9l2g2 streptomyce
10	43	54.4	134	16	Q8XXD3	Q8xxd3 ralstonia s
11	43	54.4	337	2	Q93H57	Q93h57 streptomyce
12	43	54.4	436	4	Q96S01	Q96s01 homo sapien
13	43	54.4	1070	4	O94938	O94938 homo sapien
14	43	54.4	3036	4	Q8TDJ6	Q8tdj6 homo sapien
15	42	53.2	255	5	Q9VHS9	Q9vhs9 drosophila
16	42	53.2	267	10	Q9ZP33	Q9zp33 lycopersico
17	42	53.2	363	17	Q9UYW2	Q9uyw2 pyrococcus
18	42	53.2	364	17	O58378	O58378 pyrococcus
19	41.5	52.5	567	16	Q8ZDN4	Q8zdn4 yersinia pe
20	41	51.9	104	12	Q84192	Q84192 newcastle d
21	41	51.9	160	5	Q8T2X5	Q8t2x5 trypanosoma
22	41	51.9	179	16	Q9I669	Q9i669 pseudomonas
23	41	51.9	270	16	O33286	O33286 mycobacteri
24	41	51.9	274	10	Q9SDY1	Q9sdy1 glycine max
25	41	51.9	292	16	Q93JJ4	Q93jj4 streptomyce
26	41	51.9	305	16	Q9RSH0	Q9rsh0 deinococcus
27	41	51.9	635	4	Q13476	Q13476 homo sapien
28	41	51.9	780	16	Q9A751	Q9a751 caulobacter
29	41	51.9	873	11	Q61097	Q61097 mus musculu
30	41	51.9	1548	4	Q9NYW9	Q9nyw9 homo sapien
31	41	51.9	2161	4	Q9Y566	Q9y566 homo sapien
32	40	50.6	152	10	Q9FSR2	Q9fsr2 oryza sativ
33	40	50.6	171	2	Q93S28	Q93s28 rhizobium t
34	40	50.6	187	16	Q9K3X3	Q9k3x3 streptomyce
35	40	50.6	202	4	Q96NR5	Q96nr5 homo sapien
36	40	50.6	211	16	Q9L1P1	Q9l1p1 streptomyce
37	40	50.6	212	16	Q9A4W1	Q9a4w1 caulobacter
38	40	50.6	256	15	O11932	O11932 human immun
39	40	50.6	281	2	Q53905	Q53905 streptomyce
40	40	50.6	281	16	Q93IZ2	Q93iz2 streptomyce
41	40	50.6	333	16	Q9KZW9	Q9kzw9 streptomyce
42	40	50.6	374	2	Q54527	Q54527 streptomyce
43	40	50.6	733	16	Q9X7Y4	Q9x7y4 streptomyce
44	40	50.6	789	10	Q8S1W8	Q8s1w8 oryza sativ
45	40	50.6	947	10	Q9SSA6	Q9ssa6 arabidopsis

ALIGNMENTS

RESULT 1
Q9SM14
ID Q9SM14 PRELIMINARY; PRT; 367 AA.
AC Q9SM14;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SBP-domain protein 6 (Fragment).
GN SBP6.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. T232; TISSUE=EARLY FEMALE INFLORESCENCE;
RX MEDLINE=97446501; PubMed=9301089;
RA Cardon G.H., Hoehmann S., Nettesheim K., Saedler H., Huijser P.;
RT "Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a
RT novel gene involved in the floral transition.";
RL Plant J. 12:367-377(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. T232; TISSUE=EARLY FEMALE INFLORESCENCE;
RX MEDLINE=99453765; PubMed=10524240;
RA Cardon G.H., Hoehmann S., Klein J., Nettesheim K., Saedler H.,
RA Huijser P.;
RT "Molecular characterization of the Arabidopsis SBP-box genes.";
RL Gene 237:91-104(1999).
DR EMBL; AJ011619; CAB56632.1; -.
DR InterPro; IPR004333; SBP_plant_prot.
DR Pfam; PF03110; SBP; 1.
FT NON_TER 1 1
SQ SEQUENCE 367 AA; 41043 MW; F86C6277D498F715 CRC64;

Query Match 67.1%; Score 53; DB 10; Length 367;
Best Local Similarity 76.9%; Pred. No. 1.6;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 RTAAHPAQRRPWR 13
           | || || |||||
Db     279 RRAATPAARRPWR 291


RESULT 2
Q90Z68
ID   Q90Z68      PRELIMINARY;      PRT;   454 AA.
AC   Q90Z68;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Iroquois3 homeobox protein.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21332328; PubMed=11438735;
RA   Kudoh T., Dawid I.B.;
RT   "Role of the iroquois3 homeobox gene in organizer formation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7852-7857(2001).
DR   EMBL; AF340184; AAK72232.1; -.
DR   InterPro; IPR001356; Homeobox.
DR   Pfam; PF00046; homeobox; 1.
DR   ProDom; PD000010; Homeobox; 1.
DR   PROSITE; PS00027; HOMEOBOX_1; UNKNOWN_1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
KW   DNA-binding; Homeobox; Nuclear protein.
SQ   SEQUENCE   454 AA;  50682 MW;  4C13EB8B5E1A3071 CRC64;

  Query Match             58.2%;  Score 46;  DB 13;  Length 454;
  Best Local Similarity   72.7%;  Pred. No. 23;
  Matches    8;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRP 11
           :|| || ||||
Db     428 KTAFHPVQRRP 438


RESULT 3
Q9CRM1
ID   Q9CRM1      PRELIMINARY;      PRT;   358 AA.
AC   Q9CRM1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2610001E17Rik protein (Fragment).
GN   2610001E17RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=WOLFFIAN DUCT INCLUDES SURROUNDING REGION;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK020169; BAB32018.1; -.
DR   MGD; MGI:1915146; 2610001E17Rik.
FT   NON_TER       1       1
FT   NON_TER     358     358
SQ   SEQUENCE   358 AA;  40575 MW;  8380BA079871D114 CRC64;

  Query Match             57.0%;  Score 45;  DB 11;  Length 358;
  Best Local Similarity   63.6%;  Pred. No. 27;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TAAHPAQRRPW 12
           | |:|| :|||
Db     152 TTAYPATQRPW 162


RESULT 4
Q8R2G6
ID   Q8R2G6      PRELIMINARY;      PRT;   949 AA.
AC   Q8R2G6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   URB precursor.
GN   URB.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=WHITE ADIPOSE TISSUE;
RX   MEDLINE=21670972; PubMed=11812002;
RA   Aoki K., Sun Y., Aoki S., Wada K., Wada E.;
RT   "Cloning, expression, and mapping of a gene that is upregulated in
RT   adipose tissue of mice deficient in bombesin receptor subtype-3.";
RL   Biochem. Biophys. Res. Commun. 290:1282-1288(2002).
DR   EMBL; AB075019; BAB85613.1; -.
KW   Signal.
FT   SIGNAL        1      23       POTENTIAL.
SQ   SEQUENCE   949 AA;  107640 MW;  62693C715C36F6AB CRC64;

  Query Match             57.0%;  Score 45;  DB 11;  Length 949;
  Best Local Similarity   63.6%;  Pred. No. 66;
  Matches    7;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 TAAHPAQRRPW 12
           | |:|| :|||
Db     374 TTAYPATQRPW 384


RESULT 5
O92501
ID   O92501      PRELIMINARY;      PRT;   106 AA.
AC   O92501;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AcMNPV orf149.
GN   ORF_125.
OS   Bombyx mori nuclear polyhedrosis virus (BmNPV).
OC   Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC   Nucleopolyhedrovirus.
OX   NCBI_TaxID=10458;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T3;
```

```
RX   MEDLINE=97329351; PubMed=9185864;
RA   Kamita S.G., Maeda S.;
RT   "Sequencing of the putative DNA helicase-encoding gene of the Bombyx
RT   mori nuclear polyhedrosis virus and fine-mapping of a region involved
RT   in host range expansion.";
RL   Gene 190:173-179(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T3;
RX   MEDLINE=99281911; PubMed=10355780;
RA   Gomi S., Majima K., Maeda S.;
RT   "Sequence analysis of the genome of Bombyx mori
RT   nucleopolyhedrovirus.";
RL   J. Gen. Virol. 80:1323-1337(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T3;
RA   Maeda S.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T3;
RA   Gomi S.;
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; L33180; AAC63815.1; -.
SQ   SEQUENCE   106 AA;  12314 MW;  4524364EFA9E6168 CRC64;

  Query Match             55.7%;  Score 44;  DB 12;  Length 106;
  Best Local Similarity   70.0%;  Pred. No. 12;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5 HPAQRRPWRA 14
           || || || |
Db      28 HPYQRAPWNA 37


RESULT 6
Q9ABC2
ID   Q9ABC2      PRELIMINARY;      PRT;   259 AA.
AC   Q9ABC2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Inositol monophosphatase family protein.
GN   CC0308.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005704; AAK22295.1; -.
DR   TIGR; CC0308; -.
DR   InterPro; IPR000760; Inositol_P.
DR   Pfam; PF00459; inositol_P; 1.
DR   PROSITE; PS00629; IMP_1; 1.
DR   PROSITE; PS00630; IMP_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   259 AA;  27443 MW;  C62A12659711EB91 CRC64;

  Query Match             55.7%;  Score 44;  DB 16;  Length 259;
  Best Local Similarity   77.8%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 AHPAQRRPW 12
           |||| | ||
Db     159 AHPAWREPW 167


RESULT 7
Q9RV04
ID   Q9RV04      PRELIMINARY;      PRT;   363 AA.
AC   Q9RV04;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein DR1226.
GN   DR1226.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001970; AAF10800.1; -.
DR   TIGR; DR1226; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   363 AA;  38715 MW;  FD86759A3BD5782E CRC64;

  Query Match             55.7%;  Score 44;  DB 16;  Length 363;
  Best Local Similarity   75.0%;  Pred. No. 39;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 AAHPAQRRPWRA 14
           || ||| | |||
Db     178 AALPAQARDWRA 189


RESULT 8
Q9U1K8
ID   Q9U1K8      PRELIMINARY;      PRT;  2160 AA.
AC   Q9U1K8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:BACR7C10.2 protein.
GN   EG:BACR7C10.2 OR CG10260.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ferraz C., Vidal S., Brun C., Bucheton A., Demaille J.G.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL121804; CAB65858.1; -.
DR   FlyBase; FBgn0040335; EG:BACR7C10.2.
```

```
DR   InterPro; IPR001263; PI3Ka.
DR   InterPro; IPR000403; PI3_PI4_kinase.
DR   Pfam; PF00613; PI3Ka; 1.
DR   Pfam; PF00454; PI3_PI4_kinase; 1.
DR   SMART; SM00145; PI3Ka; 1.
DR   SMART; SM00146; PI3Kc; 1.
DR   PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR   PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR   PROSITE; PS50290; PI3_4_KINASE_3; 1.
SQ   SEQUENCE   2160 AA;  242387 MW;  061F3383DCBDB8B3 CRC64;

  Query Match             55.7%;  Score 44;  DB 5;  Length 2160;
  Best Local Similarity   69.2%;  Pred. No. 2e+02;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWR 13
           | ||||  ||| |
Db     224 RAAAHPEGRRPRR 236


RESULT 9
Q9L2G2
ID   Q9L2G2      PRELIMINARY;      PRT;   326 AA.
AC   Q9L2G2;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO2521.
GN   SCO2521 OR SCC121.24C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL137166; CAB69742.1; -.
DE   Hypothetical protein.
SQ   SEQUENCE   326 AA;  35264 MW;  C7EDF0218F42B29C CRC64;

  Query Match             55.1%;  Score 43.5;  DB 16;  Length 326;
  Best Local Similarity   47.4%;  Pred. No. 42;
  Matches    9;  Conservative    2;  Mismatches    1;  Indels    7;  Gaps    1;

Qy       1 RTAAH-------PAQRRPW 12
           :|:||       | |||||
Db      95 QTSAHCRLPGSGPDQRRPW 113


RESULT 10
Q8XXD3
ID   Q8XXD3      PRELIMINARY;      PRT;   134 AA.
AC   Q8XXD3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein RSc2182.
GN   RSC2182 OR RS01417.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646068; CAD15889.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   134 AA;  14948 MW;  5CF2B7F29BDC9DF9 CRC64;

  Query Match             54.4%;  Score 43;  DB 16;  Length 134;
  Best Local Similarity   57.1%;  Pred. No. 22;
  Matches    8;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWRA 14
           |  ||  ||  |||
Db      25 RAGAHATQRTVWRA 38


RESULT 11
Q93H57
ID   Q93H57      PRELIMINARY;      PRT;   337 AA.
AC   Q93H57;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Ornithine carbamoyltransferase.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070951; BAB69334.1; -.
DR   InterPro; IPR002029; Asp/Orn_COtranf.
DR   Pfam; PF00185; OTCace; 1.
DR   Pfam; PF02729; OTCace_N; 1.
DR   PROSITE; PS00097; CARBAMOYLTRANSFERASE; UNKNOWN_1.
KW   Transferase.
SQ   SEQUENCE   337 AA;  36416 MW;  C7ED1299BBF5B019 CRC64;

  Query Match             54.4%;  Score 43;  DB 2;  Length 337;
  Best Local Similarity   57.1%;  Pred. No. 52;
  Matches    8;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWRA 14
           |||  ||: | | |
Db     133 RTAGDPAEMRAWAA 146


RESULT 12
Q96S01
ID   Q96S01      PRELIMINARY;      PRT;   436 AA.
AC   Q96S01;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 44.9 kDa protein.
```

```
GN   GS82.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21096910; PubMed=11157797;
RA   Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA   Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA   Higgs D.R.;
RT   "Sequence, structure and pathology of the fully annotated terminal 2
RT   Mb of the short arm of human chromosome 16.";
RL   Hum. Mol. Genet. 10:339-352(2001).
DR   EMBL; AE006465; AAK61265.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   436 AA;  44912 MW;  8090D76ACD7BE6E8 CRC64;

  Query Match             54.4%;  Score 43;  DB 4;  Length 436;
  Best Local Similarity   69.2%;  Pred. No. 65;
  Matches    9;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWR 13
           | ||   ||||||
Db     239 RPAALWTQRRPWR 251


RESULT 13
O94938
ID   O94938      PRELIMINARY;      PRT;  1070 AA.
AC   O94938;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   KIAA0856 protein (Fragment).
GN   KIAA0856.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99156230; PubMed=10048485;
RA   Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.
RT   XII.The complete sequences of 100 new cDNA clones from brain which
RT   code for large proteins in vitro.";
RL   DNA Res. 5:355-364(1998).
CC   -!- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
DR   EMBL; AB020663; BAA74879.1; -.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 5.
DR   SMART; SM00320; WD40; 5.
DR   PROSITE; PS50082; WD_REPEATS_2; 2.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Repeat; WD repeat.
FT   NON_TER       1       1
SQ   SEQUENCE   1070 AA;  120197 MW;  CC96C7BB01963511 CRC64;

  Query Match             54.4%;  Score 43;  DB 4;  Length 1070;
  Best Local Similarity   57.1%;  Pred. No. 1.5e+02;
  Matches    8;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWRA 14
           | ||||   | | |
Db     404 RLAAHPLNNRMWAA 417


RESULT 14
Q8TDJ6
ID   Q8TDJ6      PRELIMINARY;      PRT;  3036 AA.
AC   Q8TDJ6;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Rabconnectin.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21895900; PubMed=11809763;
RA   Nagano F., Kawabe H., Nakanishi H., Shinohara M., Deguchi-Tawarada M.,
RA   Takeuchi M., Sasaki T., Takai Y.;
RT   "Rabconnectin-3, a Novel Protein That Binds Both GDP/GTP Exchange
RT   Protein and GTPase-activating Protein for Rab3 Small G Protein
RT   Family.";
RL   J. Biol. Chem. 277:9629-9632(2002).
DR   EMBL; AF389880; AAL93215.1; -.
SQ   SEQUENCE   3036 AA;  339753 MW;  C611C9AA46D3BBFB CRC64;

  Query Match             54.4%;  Score 43;  DB 4;  Length 3036;
  Best Local Similarity   57.1%;  Pred. No. 3.9e+02;
  Matches    8;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 RTAAHPAQRRPWRA 14
           | ||||   | | |
Db    2370 RLAAHPLNNRMWAA 2383


RESULT 15
Q9VHS9
ID   Q9VHS9      PRELIMINARY;      PRT;  255 AA.
AC   Q9VHS9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   CG11698 protein.
GN   CG11698.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003678; AAF54220.1; -.
DR FlyBase; FBgn0037572; CG11698.
SQ SEQUENCE 255 AA; 28210 MW; 84331D44C060A5C0 CRC64;

Query Match 53.2%; Score 42; DB 5; Length 255;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q 2 TAAHPAQR 9
||||||||
Db 19 TAAHPAQR 26

Search completed: March 10, 2003, 17:15:15
Job time : 26.7692 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: March 10, 2003, 16:57:56 ; Search time 30.3333 Seconds
(without alignments)
57.107 Million cell updates/sec

Title: US-09-993-392-2
Perfect score: 75
Sequence: 1 KQHPCLDGSAGRN 13

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	75	100.0	614	15	AAR55799	Human betaine-GABA
2	75	100.0	614	17	AAR89481	Human betaine/GABA
3	75	100.0	1923	22	ABG21342	Novel human diagno
4	47	62.7	88	22	AAU51999	Propionibacterium
5	42	56.0	374	21	AAB18979	Amino acid sequenc
6	42	56.0	4601	22	ABB59371	Drosophila melanog
7	41	54.7	149	22	ABG08064	Novel human diagno
8	41	54.7	174	22	AAU46916	Propionibacterium
9	41	54.7	319	22	ABB62080	Drosophila melanog
10	40.5	54.0	232	21	AAG04837	Arabidopsis thalia
11	40.5	54.0	262	21	AAG04836	Arabidopsis thalia
12	40.5	54.0	329	21	AAG04835	Arabidopsis thalia
13	40	53.3	72	22	ABB15976	Human nervous syst
14	40	53.3	125	22	AAM87341	Human immune/haema
15	40	53.3	356	22	AAU42630	Propionibacterium
16	39	52.0	53	21	AAG03682	Human secreted pro
17	39	52.0	68	22	AAU40527	Propionibacterium
18	39	52.0	72	22	AAU45133	Propionibacterium
19	39	52.0	80	22	AAU64896	Propionibacterium
20	39	52.0	197	21	AAB42159	Human ORFX ORF1923
21	39	52.0	251	23	ABB48751	Listeria monocytog
22	39	52.0	254	22	AAU17241	Novel signal trans
23	39	52.0	260	22	AAU87198	Novel central nerv
24	39	52.0	312	22	AAU14605	Novel bone marrow
25	39	52.0	445	23	ABB91957	Herbicidally activ
26	39	52.0	598	21	AAG20654	Arabidopsis thalia
27	39	52.0	599	21	AAG20653	Arabidopsis thalia
28	39	52.0	751	23	ABB93984	Herbicidally activ
29	39	52.0	1028	22	AAB82249	Human insulin-resp
30	38	50.7	10	20	AAY25488	Insulin-like growt
31	38	50.7	14	21	AAY66801	T cell antigen rec
32	38	50.7	52	23	ABP05976	Human ORFX protein
33	38	50.7	54	21	AAB09676	IGFBP-4 IGF bindin
34	38	50.7	79	22	AAM84070	Human immune/haema
35	38	50.7	102	20	AAY25507	Insulin-like growt
36	38	50.7	123	21	AAB58228	Lung cancer associ
37	38	50.7	154	18	AAW09068	Banana bunchy top
38	38	50.7	154	18	AAW09072	Banana bunchy top
39	38	50.7	159	22	ABG27384	Novel human diagno
40	38	50.7	181	22	AAU27529	Human G-Protein Co
41	38	50.7	206	22	ABG03882	Novel human diagno
42	38	50.7	233	21	AAB09755	IGFBP-4 amino acid
43	38	50.7	233	21	AAY67291	Rat insulin-like g
44	38	50.7	236	13	AAR21834	Sequence of insuli
45	38	50.7	237	21	AAB09620	Insulin like growt

ALIGNMENTS

RESULT 1
AAR55799
ID AAR55799 standard; Protein; 614 AA.
XX
AC AAR55799;
XX
DT 21-MAY-1998 (first entry)
XX
DE Human betaine-GABA transporter.
XX
KW Gamma-aminobutyric acid; GABA; betaine; transporter; detection;
KW treatment; epilepsy; migraine; ischaemia; myoclonus; spasticity;
KW chronic pain; osmolyte; GAGAergic transmission; nervous system;
KW osmolarity.
XX
OS Homo sapiens.
XX
PN WO9415618-A.
XX
PD 21-JUL-1994.
XX
PF 04-JAN-1994; 94WO-US00119.
XX
PR 04-JAN-1993; 93US-0001738.
XX
PA (SYNA-) SYNAPTIC PHARM CORP.
XX
PI Borden LA, Smith KE, Weinshank RL;
XX
DR WPI; 1994-248881/30.
DR N-PSDB; AAQ66982.
XX
PT Isolated nucleic acid encoding mammalian betaine-gamma-

```
PT   amino:butyric acid transporter - useful to detect and treat
PT   abnormalities associated with transporter expression
XX
PS   Claim 35; Fig 1; 91pp; English.
XX
CC   A betaine transporter, cloned from MDCK dog kidney cells, has been
CC   isolated (Yamauchi et al., J. Biol. Chem. 267 (1):649-652).
CC   Betaine is an important osmolyte in the kidney, and possibly other
CC   organs.  This transporter was found to have higher affinity for
CC   GABA than for betaine, suggesting a role in GABAergic transmission.
CC   A related clone from a human brain cDNA library has now been
CC   isolated (AAQ66982).  Although the function of this transporter
CC   in the nervous system is not understood, it may serve to regulate
CC   both GABAergic transmission and osmolarity.
CC   Gene prods. may be used in the detection or treatment of epilepsy,
CC   migraine, ischaemia, myoclonus, spasticity or chronic pain.
XX
SQ   Sequence   614 AA;

  Query Match             100.0%;  Score 75;  DB 15;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQHPCLDGSAGRN 13
           |||||||||||||
Db     583 KQHPCLDGSAGRN 595


RESULT 2
AAR89481
ID   AAR89481 standard; Protein; 614 AA.
XX
AC   AAR89481;
XX
DT   30-MAY-1996  (first entry)
XX
DE   Human betaine/GABA transporter.
XX
KW   Betaine/GABA transporter; gamma-aminobutyric acid;
KW   neuropsychiatric disorders; human; rat; epilepsy; anxiety.
XX
OS   Homo sapiens.
XX
PN   WO9604790-A1.
XX
PD   22-FEB-1996.
XX
PF   16-AUG-1995;   95WO-US10579.

     16-AUG-1994;   94US-0291299.
XX
PA   (SYNA-) SYNAPTIC PHARM CORP.
XX
PI   Borden LA,  Smith KE,  Weinshank RL;
XX
DR   WPI; 1996-139355/14.
DR   N-PSDB; AAT16542.
XX
PT   Mammalian betaine gamma-aminobutyric acid transporter DNA - used to
PT   develop prods. for the study, diagnosis and therapy of GABA
PT   associated abnormalities, partic. neuro:psychiatric disorders.
XX
PS   Disclosure; Fig 1A-D; 191pp; English.
XX
CC   The DNA (AAT16542) encoding the human betaine/GABA transporter was
CC   isolated from a human striatum cDNA library using probes (AAT16538
CC   to AAT16541) based on a rat GABA transporter (GAT-2) cDNA.
CC   The region of rat betaine/GABA transporter encoded by the
CC   sequence given in AAT16543 corresponds to amino acids 84-139 of
CC   the human betaine/GABA transporter.
CC   Mammalian betaine gamma-aminobutyric acid transporter DNA and
CC   related prods. may be used for the study, diagnosis and therapy
CC   of GABA associated abnormalities, partic. neuropsychiatric
CC   disorders, such as epilespy and anxiety.
XX
SQ   Sequence   614 AA;

  Query Match             100.0%;  Score 75;  DB 17;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 0.00011;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 KQHPCLDGSAGRN 13
           |||||||||||||
Db     583 KQHPCLDGSAGRN 595


RESULT 3
ABG21342
ID   ABG21342 standard; Protein; 1923 AA.
XX
AC   ABG21342;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #21333.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS85529.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 51701; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ   Sequence   1923 AA;

  Query Match            100.0%;  Score 75;  DB 22;  Length 1923;
  Best Local Similarity   100.0%;  Pred. No. 0.00039;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 KQHPCLDGSAGRN 13
          |||||||||||||
Db   1664 KQHPCLDGSAGRN 1676


RESULT 4
AAU51999
ID   AAU51999 standard; Protein; 88 AA.
XX
AC   AAU51999;
XX
DT   27-FEB-2002  (first entry)
XX
[redacted]   Propionibacterium acnes immunogenic protein #12895.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59553.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 13194; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ   Sequence   88 AA;

  Query Match            62.7%;  Score 47;  DB 22;  Length 88;
  Best Local Similarity   75.0%;  Pred. No. 1.2;
  Matches    9;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 KQHPCLDGSAGR 12
          | || ||||| |
Db     48 KPHPVLDGSASR 59


RESULT 5
AAB18979
ID   AAB18979 standard; Protein; 374 AA.
XX
AC   AAB18979;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of a human transmembrane protein.
XX
KW   Human; transmembrane protein; cell proliferation disorder;  myeloma;
KW   reproductive disorder; smooth muscle disorder; neurological disorder;
KW   arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;
KW   allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;
KW   Alzheimer's disease; Tourette's disorder.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          49..70
FT                   /note= "leucine zipper"
FT   Modified-site   10
FT                   /note= "potential phosphorylation site"
FT   Modified-site   33
FT                   /note= "potential phosphorylation site"
FT   Modified-site   57
FT                   /note= "potential phosphorylation site"
FT   Modified-site   59
FT                   /note= "potential phosphorylation site"
FT   Modified-site   74
FT                   /note= "potential phosphorylation site"
FT   Modified-site   130
FT                   /note= "potential phosphorylation site"
FT   Modified-site   193
FT                   /note= "potential phosphorylation site"
FT   Modified-site   312
FT                   /note= "potential phosphorylation site"
FT   Modified-site   322
FT                   /note= "potential phosphorylation site"
FT   Modified-site   320
FT                   /note= "potential glycosylation site"
XX
PN   WO200056891-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07817.
XX
PR   22-MAR-1999;   99US-0125537.
PR   16-JUN-1999;   99US-0139565.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Yue H,  Lal P,  Tang YT,  Hillman JL,  Reddy R,  Bandman O,  Baughn MR;
PI   Lu DAM,  Azimzai Y,  Yang J;
XX
DR   WPI; 2000-579485/54.
DR   N-PSDB; AAA96492.
XX
PT   New human transmembrane proteins are used to treat a disease or
PT   condition associated with decreased expression of functional HTMP e.g.
PT   Tourette's disorder, angina and leukaemia -
```

```
XX
PS   Claim 1; Page 99-100; 130pp; English.
XX
CC   The present sequence represents a human transmembrane proteins (HTMP).
CC   Agonists and antagonists of the protein are used to treat a disease
CC   or condition associated with overexpression of the protein. Diseases
CC   and conditions which can be treated include cell proliferative,
CC   immunological, reproductive, smooth muscle and neurological disorders
CC   e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency
CC   syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,
CC   stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The
CC   polynucleotides may be used to detect and quantify gene expression in
CC   biopsied tissues where protein expression may be correlated with disease
CC   e.g. to determine absence, presence or excess expression of HTMP or to
CC   monitor regulation of HTMP expression during therapeutic intervention.
XX
SQ   Sequence   374 AA;

  Query Match             56.0%;  Score 42;  DB 21;  Length 374;
  Best Local Similarity   72.7%;  Pred. No. 41;
  Matches    8;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 KQHPCLDGSAG 11
           |||| |||  |
Db     114 KQHPLLDGVDG 124


RESULT 6
ABB59371
ID   ABB59371 standard; Protein; 4601 AA.
XX
AC   ABB59371;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 4905.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
     23-MAR-2000; 2000US-191637P.
     11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03474.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 4905; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   4601 AA;

  Query Match             56.0%;  Score 42;  DB 22;  Length 4601;
  Best Local Similarity   63.6%;  Pred. No. 6e+02;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 HPCLDGSAGRN 13
           ||| | :|| |
Db     439 HPCRDNNAGCN 449


RESULT 7
ABG08064
ID   ABG08064 standard; Protein; 149 AA.
XX
AC   ABG08064;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #8055.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS72251.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 38423; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence   149 AA;

  Query Match             54.7%;  Score 41;  DB 22;  Length 149;
  Best Local Similarity   63.6%;  Pred. No. 23;
  Matches    7;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 HPCLDGSAGRN 13
          |||| |:  ||
Db     69 HPCLGGAPVRN 79


RESULT 8
AAU46916
ID   AAU46916 standard; Protein; 174 AA.
XX
AC   AAU46916;
XX
DT   27-FEB-2002  (first entry)

     Propionibacterium acnes immunogenic protein #7812.

KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59535.
XX

     Propionibacterium acnes polypeptides and nucleic acids useful for
     vaccinating against and diagnosing infections, especially useful for
     treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 8111; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   174 AA;

  Query Match             54.7%;  Score 41;  DB 22;  Length 174;
  Best Local Similarity   54.5%;  Pred. No. 27;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 HPCLDGSAGRN 13
          |||::  || |
Db     77 HPCVERQAGEN 87


RESULT 9
ABB62080
ID   ABB62080 standard; Protein; 319 AA.
XX
AC   ABB62080;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 13032.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06183.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 13032; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   319 AA;

  Query Match             54.7%;  Score 41;  DB 22;  Length 319;
  Best Local Similarity   58.3%;  Pred. No. 52;
  Matches    7;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 KQHPCLDGSAGR 12
          ::| | ||| ||
Db    249 RRHLCTDGSNGR 260


RESULT 10
```

AAG04837
ID AAG04837 standard; Protein; 232 AA.
XX
AC AAG04837;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1014.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
[redacted] 25-FEB-2000; 2000EP-0301439.

[redacted] 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
[redacted] 18-MAY-1999; 99US-0134768.
[redacted] 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 54.0%; Score 40.5; DB 21; Length 232;
Best Local Similarity 40.9%; Pred. No. 45;
Matches 9; Conservative 2; Mismatches 2; Indels 9; Gaps 1;

```
Qy       1 KQHP---------CLDGSAGRN 13
             ||||         || |: |:|
Db     141 KQHPEAAWDILMYCLCGAVGQN 162
```

RESULT 11
AAG04836
ID AAG04836 standard; Protein; 262 AA.
XX
AC AAG04836;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1013.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 54.0%; Score 40.5; DB 21; Length 262;
Best Local Similarity 40.9%; Pred. No. 51;
Matches 9; Conservative 2; Mismatches 2; Indels 9; Gaps 1;

```
Qy       1 KQHP---------CLDGSAGRN 13
             ||||         || |: |:|
Db     171 KQHPEAAWDILMYCLCGAVGQN 192
```

RESULT 12
AAG04835
ID AAG04835 standard; Protein; 329 AA.
XX
AC AAG04835;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1012.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 54.0%; Score 40.5; DB 21; Length 329;
Best Local Similarity 40.9%; Pred. No. 65;
Matches 9; Conservative 2; Mismatches 2; Indels 9; Gaps 1;

```
Qy       1 KQHP---------CLDGSAGRN 13
             ||||         || |: |:|
         238 KQHPEAAWDILMYCLCGAVGQN 259
```

RESULT 13
ABB15976
ID ABB15976 standard; Protein; 72 AA.
XX
AC ABB15976;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polypeptide SEQ ID NO 4633.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
DR N-PSDB; ABA12302.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Claim 11; SEQ ID NO 4633; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 72 AA;

Query Match 53.3%; Score 40; DB 22; Length 72;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 HPCLDGSA 10
||||:| |
Db 34 HPCLEGQA 41

RESULT 14
AAM87341
ID AAM87341 standard; Protein; 125 AA.
XX
AC AAM87341;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:14934.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
DR N-PSDB; AAK60122.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Claim 11; SEQ ID NO 14934; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting

CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 125 AA;

Query Match 53.3%; Score 40; DB 22; Length 125;
Best Local Similarity 41.7%; Pred. No. 28;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KQHPCLDGSAGR 12
:| ||: |::|:
Db 47 RQQPCMGGASGK 58

RESULT 15
[redacted]42630
AAU42630 standard; Protein; 356 AA.
XX
AC AAU42630;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #3526.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
[redacted] Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59518.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 3825; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 356 AA;

Query Match 53.3%; Score 40; DB 22; Length 356;
Best Local Similarity 50.0%; Pred. No. 87;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KQHPCLDGSAGR 12
:|||| ::||
Db 166 RQHPCRAAASGR 177

Search completed: March 10, 2003, 17:13:13
Job time : 31.3333 secs